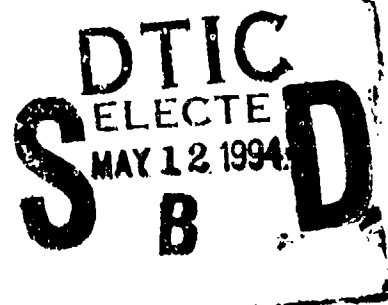
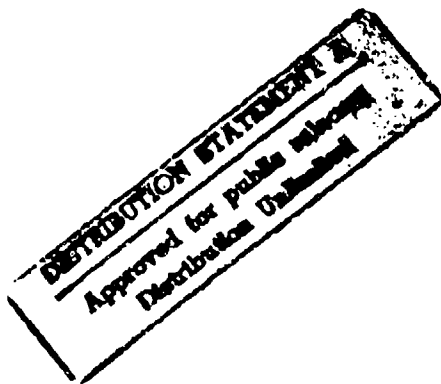


RESEARCH PAPER
CAA-RP-93-4

2

**PERSONNEL ATTRITION RATES IN
HISTORICAL LAND COMBAT OPERATION
A CATALOG OF ATTRITION AND CASUAL
DATA BASES ON DISKETTES USABLE
WITH PERSONAL COMPUTERS**

SEPTEMBER 1993



PREPARED BY
OFFICE, SPECIAL ASSISTANT FOR MODEL VALIDATION

US ARMY CONCEPTS ANALYSIS AGENCY
3120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

94-14111



07 5 10 047

DISCLAIMER

The findings of this report are not to be construed as an official Department of the Army position, policy, or decision unless so designated by other official documentation. Comments or suggestions should be addressed to:

**Director
US Army Concepts Analysis Agency
ATTN: CSCA-MV
8120 Woodmont Avenue
Bethesda, MD 20814-2797**

REPORT DOCUMENTATION PAGE			Form Approved OPM No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources gathering and maintaining the data needed, and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE September 1993		3. REPORT TYPE AND DATES COVERED March 1993-September 1993
4. TITLE AND SUBTITLE Personnel Attrition Rates in Historical Land Combat Operations: A Catalog of Attrition and Casualty Data Bases on Diskettes Usable with Personal Computers			5. FUNDING NUMBERS	
6. AUTHOR(S) Dr. Robert L. Hemibold				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Concepts Analysis Agency 8120 Woodmont Avenue Bethesda, MD 20814-2797			8. PERFORMING ORGANIZATION REPORT NUMBER CAA-RP-93-4	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE A	
13. ABSTRACT (Maximum 200 words) This paper provides a catalog and guide to the principal data bases on personnel casualties and attrition that are available on diskette for use with personal computers. It will be of considerable value to all who use personnel casualty and attrition data in studies and analyses, weapons evaluation, wargames and simulations, model validation, assessing the utility of protective measures, and so forth.				
14. SUBJECT TERMS Personnel, combat, attrition, battle, casualties, losses			15. NUMBER OF PAGES 177	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

(THIS PAGE INTENTIONALLY LEFT BLANK)

RESEARCH PAPER
CAA-RP-93-4

AD-A279069

**PERSONNEL ATTRITION RATES IN HISTORICAL
LAND COMBAT OPERATIONS:
A CATALOG OF ATTRITION AND CASUALTY DATA BASES
ON DISKETTES USABLE WITH PERSONAL COMPUTERS**

September 1993

Prepared by
Dr. Robert L. Helmbold
OFFICE, SPECIAL ASSISTANT FOR MODEL VALIDATION
US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20814-2797

PREFACE

The Personnel Attrition Rates (PAR) Study as a whole is limited to studying personnel strengths and battle casualties in historical land combat operations. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope, as are personnel losses in models, simulations, wargames, field experiments, or training exercises (like those of the National Training Center).

Phase 1, or PAR-P1, was devoted to assembling the available data and past studies on personnel strengths and attrition rates in land combat operations, preparing a comprehensive bibliography of it, and planning the approach to subsequent phases. Its results were published as "Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, June 1993 (AD-A268 787). Its collection of data and data-based studies consists of the files of pertinent documents maintained at the US Army Concepts Analysis Agency.

Phase 2, or PAR-P2, is devoted primarily to computerizing selected portions of the available data and to compiling a catalog of computerized data on personnel casualties and attrition usable on personal computers. That is the substance of this paper, which will be useful to military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of wargames and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

PAR-P2 also includes performing selected analyses. The only such analysis performed to date was reported as "Personnel Attrition Rates in Land Combat Operations: Susceptibility and Vulnerability of Major Anatomical Regions," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-3, August 1993 (AD-A270 766).



REPLY TO
ATTENTION OF:

DEPARTMENT OF THE ARMY

US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797



12 APR 1994

CSCA-MV

MEMORANDUM FOR Deputy Under Secretary of the Army (OR), Headquarters,
Department of the Army, Washington, DC 20310

SUBJECT: Personnel Attrition Rates in Historical Land Combat Operations: A
Catalog of Attrition and Casualty Data Bases on Diskettes Usable With Personal
Computers

1. The US Army Concepts Analysis Agency (CAA) is pleased to publish this Research Paper by Dr. Robert L. Helmbold. Its catalog of data bases on diskettes suitable for use with personal computers should be useful to military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of wargames and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

2. Questions or inquiries should be directed to the Office of the Special Assistant for Model Validation, U. S. Army Concepts Analysis Agency (CSCA-MV), 8120 Woodmont Avenue, Bethesda, MD 20814-2797, (301) 295-1611 or DSN 295-1611

E. B. VANDIVER III
Director

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution	
Availability Codes	
Dist	Avail and/or Special
A-1	

(THIS PAGE INTENTIONALLY LEFT BLANK)

CAA	PERSONNEL ATTRITION RATES IN HISTORICAL LAND COMBAT OPERATIONS: A CATALOG OF ATTRITION AND CASUALTY DATA BASES ON DISKETTES USABLE WITH PERSONAL COMPUTERS	SUMMARY CAA-RP-93-4
------------	---	--------------------------------

THE REASON FOR PERFORMING THIS STUDY was that a catalog of computerized data bases on personnel casualties and attrition is needed. Such a catalog will be of great value to all who use personnel casualty and attrition data in studies and analyses, weapons evaluation, wargames and simulations, model validation, assessing the utility of protective measures, and so forth.

THE STUDY SPONSOR was the Director, US Army Concepts Analysis Agency (CAA).

THE STUDY OBJECTIVE was to provide the Army with a convenient catalog of the readily-available computerized data bases on personnel casualties and attrition. A coordinate objective is to provide diskette copies of these data bases to the Defense Technical Information Center (DTIC) for archival storage.

THE SCOPE OF THE STUDY is limited to a selected class of data base. The criteria for inclusion of a data base are as follows (roughly in order of importance). The data base must be:

(a) In the public domain, so that copies can be made available to governmental agencies and others without restriction and for (at worst) a nominal cost. However, for the sake of completeness, some important proprietary data bases can be described, even if their data cannot be made available through DTIC.

(b) In data base form (i.e., consist primarily of tabulations rather than narratives).

(c) Such as to contain information on military operations in and/or outcomes of battles or wars.

(d) Available on diskettes usable with personal computers.

(e) Useful to many military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of wargames and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

(f) Difficult or inconvenient for individuals and separate study teams to generate or recreate, but which would be more frequently used if they were readily available through DTIC.

THE MAIN ASSUMPTION of this paper is that the bulk of the pertinent data are on file at CAA.

THE BASIC APPROACH used in this study was to adopt existing data bases or to prepare new ones as needed, and to catalog them.

THE PRINCIPAL FINDINGS of this work are that several data bases are available for use with personal computers. Such data bases can be used for a variety of analyses, including model validation, projections of personnel casualty and attrition rates for use in studies and analyses, weapons evaluation, wargames and simulations, assessing the utility of protective measures, and so forth.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Scenarios and Model Validation Division.

COMMENTS AND SUGGESTIONS may be sent to the Director, US Army Concepts Analysis Agency, ATTN: CSCA-MV, 8120 Woodmont Avenue, Bethesda, Maryland, 20814-2797.

(THIS PAGE INTENTIONALLY LEFT BLANK)

CONTENTS

PREFACE

CHAPTER

Page

1	EXECUTIVE SUMMARY	1-1
	Background	1-1
	Objective	1-1
	Scope	1-1
	Assumptions	1-3
	Approach	1-3
	Observations	1-3
2	OBSERVATIONS ON VARIOUS ASPECTS OF DATA BASE CONSTRUCTION	2-1
	Introduction	2-1
	Types and Uses of Historical Data	2-1
	Kinds of Data that Should be Recorded	2-3
	Suggestions on Manner of Presentation	2-6
	Observations on Accuracy Requirements	2-11
	Dealing With Degrees of Confidence	2-16
	Quality Control Measures	2-18
	Additional Data Needs	2-19
	Suitability of Personnel Skills and Characteristics	2-21
	Summary of Main Points	2-22

APPENDIX

A	List of Data Bases	A-1
B	Catalog of Data Bases	B-1
C	The ICPSR Data Bases	C-1
D	Other Data Bases or Sources of Information	D-1
E	Distribution	E-1

GLOSSARY	Glossary-1
-----------------------	------------

(THIS PAGE INTENTIONALLY LEFT BLANK)

CHAPTER 1

EXECUTIVE SUMMARY

1-1. BACKGROUND. In April 1992, the US Army Concepts Analysis Agency (CAA) started a three-phased study of Personnel Attrition Rates (PAR). The first phase involved collecting as much as possible of the available documented information on personnel attrition rates in land combat operations. Its collection of data and data-based studies consists of the files of pertinent documents maintained at the US Army Concepts Analysis Agency. This collection is described in the paper "Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, June 1993 (AD-A268 787).

Phase 2 of PAR, abbreviated as PAR-P2, is devoted primarily to computerizing selected portions of the available data and to compiling a catalog of computerized data on personnel casualties and attrition usable on personal computers. That is the substance of this paper, which will be useful to many military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of wargames and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

PAR-P2 also includes performing selected analyses. The only such analysis performed to date was reported as "Personnel Attrition Rates in Land Combat Operations: Susceptibility and Vulnerability of Major Anatomical Regions," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-3, August 1993 (AD-A270 766).

1-2. OBJECTIVE. The main reason for performing this study was to make available to a wider audience the catalog of data bases on personnel casualties and attrition prepared during the PAR studies.

1-3. SCOPE. PAR as a whole is limited to studying personnel strengths and battle casualties of land combat forces. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope. PAR is concerned only with historical data on actual combat operations: it does *not* deal with personnel losses in models, simulations, wargames, field experiments, or training exercises (like those of the National Training Center). PAR focuses mainly on either original or translated works in English, although some important work in other languages may be included. Studies of personnel attrition are also included, provided they contain cogent analyses of a publicly available, nonproprietary body of tabulated data on attrition in actual combat operations. Since trends in attrition

over long periods of time are of interest, data on ancient as well as recent battles are solicited. However, as no contract support is anticipated and in-house resources are limited, no systematic effort is made to extract data from the archives or primary source materials, and no original historical research is envisioned. Thus, PAR relies almost exclusively on secondary works that contain data in readily usable tabulated form. All works received prior to the cutoff date of 31 May 1993 are included in the final report on Phase 1, "Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, June 1993.

The scope of the present paper is limited primarily to presenting a catalog of data bases of personnel casualties and attrition that are currently available on diskettes suitable for use with personal computers. The chief criteria for inclusion of a data base are as follows (roughly in order of importance). The data base should be:

- a. In the public domain, so that copies can be made available to governmental agencies and others without restriction and for (at worst) a nominal cost. However, for the sake of completeness, some important proprietary data bases can be described, even if their data cannot be made available through the Defense Technical Information Center (DTIC).
- b. In data base form (i.e., consist primarily of tabulations rather than narratives).
- c. Such as to contain information on military operations in and/or outcomes of battles or wars.
- d. Available on diskettes usable with personal computers. (Some of the data bases in document form collected during PAR-P1 may be converted to digital form during PAR-P2 if that appears to offer a significant benefit to subsequent phases of the PAR study. However, the extent of such digitization will be drastically limited by available time and effort.)
- e. Useful to many military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of wargames and analogous combat simulations; military historians; students of military art and science; and others with similar interests.
- f. Difficult or inconvenient for individuals and separate study teams to generate or recreate, but which would be more frequently used if readily available through DTIC.

In some cases, copyright or other restrictions prevent making the actual data available through DTIC. In such cases, we can describe the data base and its availability, but cannot make the actual data available for general use. In addition, some notable data bases that are in principle available are also included, even though they are not available on diskettes suitable for use with personal computers, because it seemed inappropriate to artificially exclude the few data bases of this sort that came to our attention. However, no systematic attempt was made to seek them out.

1-4. ASSUMPTIONS. The main assumption of this paper is that the bulk of the pertinent works have been collected and are on file at CAA.

1-5. APPROACH. The basic approach used in this study was to describe each of the available data bases in a common format. This format allows for inclusion of information on the following important data base characteristics:

- a. General (full bibliographic reference to the primary documentation, description of the kinds of information it provides, other important factual information, and a statement of the situational descriptors used).
- b. Data Sources Used.
- c. Diskette Format (computer hardware and software compatibility restrictions, file descriptions, data field specifications, and so forth).
- d. Other and Miscellaneous (examples of the use of this data base, other informative remarks).
- e. Comments and Critique (discussion of the strong and weak points of this data base).

1-6. OBSERVATIONS

a. Appendix A lists the data bases in alphabetical order according to a short title adopted for ease of reference. Thus, ACSDB-1990 refers to the Ardennes Combat Simulation Data Base, which was current or originally created in the year 1990. Appendix B contains the detailed catalog of information on each data base, covering the items of information mentioned in paragraph 1-5, above. Appendix C is devoted to a particular class of proprietary data base, namely, those maintained and provided by the Inter-University Consortium for Political and Social Research (ICPSR). Appendix D mentions some additional important data bases and sources of information that were omitted from Appendix B's catalog of data bases for one reason or another.

b. The reader is warned that the data bases listed in Appendix B often partially overlap. That is, the same operation or action may be listed in more than one of them. One consequence of this is that different data bases may give different values for a given operation or action. For example, the values for the US Civil War battle of Antietam can be used to construct the following comparison, where TBC stands for "Total Battle Casualties," i.e., for the sum of the killed in action (KIA), wounded in action (WIA), and captured or missing in action (CMIA):

<u>Data Base</u>	<u>USA Strength</u>	<u>USA TBC</u>	<u>CSA Strength</u>	<u>CSA TBC</u>
BWSH-1993	60,000	12,500	40,000	11,200
CDB90-1991	90,000	12,410	46,000	13,700
LIVRMORE-1993	75,316	12,410	51,844	13,724
PARMISC-1993	75,000	12,140	41,000	13,724
<u>SP128-1961</u>	70,000	15,029 [†]	39,200	11,305 [†]

[†] These values include killed and wounded only.

The differences exhibited here are typical of the discrepancies between various sources.

c. Chapter 2 offers a number of observations on data base design and construction, drawn from our experiences as well as those of others.

CHAPTER 2

OBSERVATION ON VARIOUS ASPECTS OF DATA BASE CONSTRUCTION

2-1. INTRODUCTION. In this chapter we offer some observations on the construction of data bases on historical land combat operations. Throughout this chapter, we are primarily concerned with historical data bases assembled for the purpose of quantitative or scientific analysis of military operations. The presentation draws on the many years of experience we have had in the design, construction, and use of such data bases, as well as on the experiences of others. In order to give our observations some structure, we have arranged them under the following general headings:

- What are the main types and use of historical data on land combat operations?
- What data should be recorded?
- How should the data be presented?
- What accuracy is needed to support subsequent analyses?
- Should degrees of confidence in the data be estimated? If so, how should they be presented?
- What quality control measures are appropriate?
- In what areas would additional data be most useful?
- Who should be involved?

Each of these will be taken up in the following paragraphs. The final paragraph offers a concise summary of what we consider to be the main points. Much of what we say here is not new, and in many cases is little more than "obvious common sense." In particular, we have repeated (either in their words or ours) many of the observations made by McQuie, *et al*, "Multivariate Analysis of Combat," PRC R-1143, Planning Research Corporation, 1968, UNCLASSIFIED, available from DTIC as AD-643 294. In this chapter, all references to McQuie are to this work. Although our observations may be neither new nor very deep, they bear repeating since they are more honored in the breach than in the observance.

2-2. TYPES AND USES OF HISTORICAL DATA. Here we offer some observations on the main types and uses of historical data, in order to place the quantitative or scientific type of historical combat data base in perspective relative to other types and uses. Data bases may draw upon primary, secondary, or tertiary sources. Quantitative analysis uses historical data in ways that can be distinguished from most past uses.

a. **Sources.** The sources of military history are often divided into various classes depending on how proximate the records were to the actual event. The following potential sources of data can be mentioned, beginning with the "primary" or original records. In principle, historical data bases may be drawn from any one of the following sources, or from any combination of them. As a practical matter, however, most of them lean heavily on tertiary sources. As explained in more detail in paragraph 2-5, primary records, despite their immediacy and proximity to the action, are not necessarily the most accurate or complete, especially with regard to the quantitative aspects of battle. Hence, they are not always the best suited to scientific analysis.

(1) **Primary Records.** Primary records are usually those made at the time of the event. They include films, video tapes, photographs, *etc.* They also include official records made at the time, such as after-action reports, situation reports, copies of messages sent at the time of the action, unit diaries, log books, map overlays, computerized records, damage surveys, *etc.* Also included are unofficial records made at the time, such as eyewitness accounts, interviews, reportage, personal diaries, paintings and sketches, *etc.*

(2) **Secondary Sources.** These are accounts developed from the primary records and include unit histories, official histories (such as the *History of the US Army in World War II* or "Green Book" series). We also count personal memoirs, oral histories, and archaeological explorations (such as those done by the Israelis, and by the US at Little Big Horn and Fort Necessity) as secondary sources.

(3) **Tertiary Sources.** These are accounts developed from primary or secondary sources. In this category we include libraries, museums, most kinds of nonofficial history, dictionaries of battles, and previously prepared data bases.

b. **Uses of Military History.** It is useful to distinguish at least the following uses of military history.

(1) **Entertainment.** Military history is a rich source of material for human drama, vicarious adventures, thrilling accounts of human triumph or disaster, and stirring tales of individual heroism in the face of adversity. J. F. C. Fuller has called this the "bloody romance" school of military history.

(2) **Propaganda.** Military history also plays an important role by providing inspiring models of heroism, self-sacrifice, and steadfastness in war. It can profitably be used to build a sense of national solidarity through a common heritage, or to build unit morale and esprit through accounts of past triumphs. But it can also be exploited for personal vanity and/or aggrandizement, to support a particular biased thesis or policy, to intimidate potential or actual enemies, to assert peacetime leadership of a coalition based on claimed past military contributions to an alliance, and so forth.

(3) **Describe Accurately Particular Cases.** Careful and objective descriptions of particular cases satisfy our intellectual curiosity, furnish useful case studies for teaching military art and science,

and generate examples and counterexamples that illustrate or disprove claims. Such examples can provide "existence proofs" that disprove (usually exaggerated) claims that some action "never" or "always" leads to a particular result. These descriptions can also be used to delineate a range of plausible outcomes as a check on the "reasonableness" of planning factors (for instance, planning for an average personnel loss rate of 5 percent per day sustained over a period of a month is probably unrealistic, because historically such loss rates have practically never been sustained for more than a few days.)

(4) Systematic Quantitative Analysis Emphasizing Scientific and Statistical Methods.

Objective military history can provide data needed for quantitative analyses of combat through the use of descriptive statistics, tests of judiciously selected statistical hypotheses, checks on the validity of alternative scientific theories, and identifications and/or estimates of the influence of various factors contributing to battle outcomes (such as victory, rate of advance, *etc.*).

2-3. KINDS OF DATA THAT SHOULD BE RECORDED. Here we offer some observations on the kinds of data that should be included in historical data bases of land combat operations. Lacking an adequate theoretical basis, the selection of data to record is often *ad hoc*. Some guidance can be provided by erecting specific hypotheses to be tested, and then assembling data suited to such tests. Careful plans for data collection will get the best results from a given level of resources.

a. It is generally conceded that the kinds of data needed for quantitative analyses are not readily available from conventional military histories (*i.e.*, narrative accounts of the action as exemplified by the "Green Books" on the US Army in World War II). Indeed, McQuie observes that "Military history of the conventional type, and the type of information required for statistical analysis, often conflict with one another." This is one aspect of the more general statement that we have no theory of military combat adequate to guide data collection, provide an unambiguous definition of terms, *etc.* However, we can erect various hypotheses to test, and use *them* to guide the data collection. This is, in essence, a recommendation to use methods akin to those known as "strong inference," as described by Platt, John R., "Strong Inference," *Science*, v 146, n 3642, 16 October 1964, pp 347-353 and Chamberlin, T. C., "The Method of Multiple Working Hypotheses," *Science*, v 148 (1965), pp 754-759.

b. In view of our lack of a solid theoretical basis, we must expect a considerable degree of redundancy in any data base of battles. This expresses itself in correlations among the variables (the higher the correlation, the higher the redundancy). The presence of widespread redundancy or correlation among the variables complicates the statistical analysis, but seems to be unavoidable with our present state of knowledge. Without an adequate theory of combat, it is difficult to say which data are relevant and which are not. Accordingly, it is to be expected that we will inadvertently collect partially duplicative data on many variables in the data base.

c. It should be recognized that the collection of accurate statistical data is an arduous task. It cannot be done by participants, cannot be done from a distance, cannot be done later, and will not succeed without a well-conceived plan. As noted in McQuie, "The collection of accurate statistical data is a laborious and time-consuming task, which should be done in accordance with a well-defined plan of analysis. It is not something that can be done by those involved in fighting a war. Even in field experiments, it has been found to be a cardinal principle that the test subjects cannot be expected to collect experimental data that are accurate, complete, or objective. Attempts to do so have almost invariably failed miserably. This suggests that the Army institute a well-conceived plan to have independent data collectors in all future combat actions." Nor is it possible to collect data after the fact—as shown by the failure of the efforts to do so for the Vietnam War and Operation DESERT STORM. McQuie astutely predicted that "If such designed data collection plans are not initiated, it is quite possible that the statistical data inadequacies of World War II and Korea may be repeated and many valuable lessons which could be learned from Vietnam will be lost," and this, of course, is exactly what actually occurred.

d. The US Army should prepare in peacetime a careful plan for collecting and analyzing objective, accurate, detailed, comprehensive, consistent scientific data during any future combat operation. The prototypes are the operational data collection and analysis work done by British and US operations analysts during World War II and Korea (respectively). Historically, detailed data has seldom been collected and analyzed on military operations as they occurred. One of the reasons for this is that excessive reliance is often placed on data collected by the participants themselves, even though innumerable past experiences with data collection in war—and even in the more benign environment of field experiments—amply demonstrate that this is unsatisfactory. Objective, accurate and complete data cannot be obtained either by warriors properly preoccupied with fighting the battle, or at a distance from the battlefield, or at a later time. Instead, it must be done on the spot by suitably-trained independent data collectors. Their efforts need to be institutionalized to plan for a sizable effort, promote objectivity, standardize precise terminology, select and train the personnel, arrange for safeguarding and transmitting the data, and provide for systematically testing and improving the data collection plans. The importance of this is underscored by a remark of General O. P. Weyland's regarding World War II to the effect that what happened was not clearly perceived, or if clearly perceived, was not remembered, or if remembered, was not written down, or if written down was not disseminated, or if disseminated, was not read, or if read, was not understood. As expressed by McQuie, "Despite the voluminous records on file, data available on World War II and Korea are inadequate, incomplete, and probably biased." In illustration, he points out that "Unit histories and after-action reports tend to be less than objective. This became especially clear during the data search when an effort was made to reconstruct a particular battle by reviewing the after-action reports starting at the battalion level and then attempting to correlate the data from the three battalions with regimental reports. The three regimental reports in

turn bore little resemblance to the division account." The Vietnam data are no better. Uhorchak, John M., *et al*, "Final Report: Casualty Data Assessment Team Operation DESERT STORM," Letterman Army Institute of Research, Division of Military Trauma Research, Presidio of San Francisco, California 94129, Institute Report No. 469, January 1992, 115 pp, UNCLASSIFIED, observes that the original Casualty Data Assessment Team (CDAT) plan was to send a large team to Southwest Asia to survey the wounded soldiers as far forward as possible. Instead, permission was granted only to interview wounded being processed through Army general hospitals in Germany. Accordingly, the team was located too far to the rear to obtain detailed information, which as a result is now irretrievable.

Uhorchak notes that this same bureaucratic inertia and indifference ruined attempts to provide prompt data collection and analysis in every war since the US Civil War, and remarks that "One would expect that [the largest health care system in the world] ... would have a prospective and near real-time analysis of casualties as the military situation and the latest data acquisition technology might allow. ... [Yet] the Army Medical Department currently does not have a standing system to analyze combat casualties. ... We reiterate that our concept is not new, but unfortunately many of the problems we encountered also are not new." Indeed, after recalling a World War II report of The Office of The Surgeon General noting instances of badly misguided resistance to clinical and laboratory investigations in a theater of war, Uhorchak caustically remarks that "This attitude still prevails in some circles."

The efficiency of data collection can be greatly aided by a suitable plan. For example, each "pass" through the original or primary data sources is very labor intensive and therefore very time-consuming and costly. Accordingly, making repeated passes to collect a few additional facets of information, or to check some of them, is prohibitively wasteful and inefficient. Hence, it is important to have complete and thorough preparations and plans to extract the maximum information per pass, while maintaining accuracy and objectivity. These plans need to establish good, formal definitions of all the terms (even for such commonly used terms as "battle," "strength," and "losses"). It is prudent to subject the data collection plan to a preliminary trial or dress rehearsal to validate the definitions, procedures, reporting forms, and to give the data collectors a practice run before standardizing the particular scheme to use for the remainder of the data collection effort.

It is extremely important that, for each data entry, the data base include an indication of exactly how it was obtained (in sufficient detail for a subsequent investigator to reproduce the data value). For example, was it extracted directly from some source, and if so, from what source (specifying its page or table number)? If not extracted directly from some source, how was the value obtained (*i.e.*, what information is it based on, and exactly how was this information used to generate the value inserted in the data base)?

It is also important that the data base provide an indication of the degree of confidence in the accuracy of each item of quantitative data, such as a specification of its "error tolerance."

It is certainly true that our data needs have changed with the advancement of data collection and processing technology. In particular, this has given rise both to the need for large computerized data bases and to effective ways of using them. The salient characteristics of a properly organized data base are that if a variable is measured for one battle, it must be measured for all battles, and if measured for one side must also be measured for the other. It is also essential that the information be organized in a way that facilitates its easy retrieval, manipulation, and use in scientific analyses.

In order to assure the efficient attainment of these objectives, a systematic data collection plan for future combat engagements is required. The computerized data bases of battles and other militarily related activities listed in the appendixes can be consulted for excellent suggestions on what to plan for. Perhaps identifying the kinds of data that should be recorded can also be aided by considering the following general categories, which were developed in connection with the DESERT STORM operation:

- **Environmental Data**—data that are directly available from and dependent on the physical environment.
- **Functional Area Data**—refers primarily to the support functions that must be performed in both the peacetime and wartime environment. In each case, the data collected should be in reference to both the demand for support and the capability of the respective elements to provide support. Also of concern is how demand and capability vary over time.
- **Operational Data**—refers to the political, strategic, operational, and tactical planning, and implementation as it evolves to meet changing circumstances.
- **Performance Data**—refers to the performance of people and systems as affected specifically by the local operational environment.
- **Force Definition Data**—refers to the location, composition, and status of all forces in theater, to include US forces, allies, and (to the extent known) threat.

Each of these general categories can be further considered in relation to the following operational phases:

- Strategic and Intratheater Deployment.
- Peacetime/Prewar Preparation.
- Employment/War Initiation.
- War Sustainment.
- War Termination.
- Postwar Conditions.

2.4. SUGGESTIONS ON MANNER OF PRESENTATION. Here we offer some observations on how historical data bases on land combat operations can be presented. Systematically organized combat data bases, with the same data items provided for each of a large number of cases, with all data developed on a common basis for comparability and compatibility, are needed. These data can easily be manipulated for analysis purposes if provided in computerized format. Hard copy versions could be prepared directly from the computerized versions.

a. **Data Base Form.** Since we are concerned here with the use of military history for systematic quantitative analysis employing scientific and statistical methods, systematically organized data bases are essential. This is because data bases provide detailed data, systematically arranged, with the same data items provided for each of a fairly large number of battles, and with all the data developed on a common basis for comparability and compatibility. This has been eloquently expressed by McQuie as follows: "A meaningful analysis of anything complicated requires that its characteristics be studied in relation to each other. For combat, such an analysis requires that identical data be accumulated from one battle to the next, and the same type of data are required about enemy forces as about our own."

b. **Digital Format.** How should data bases be presented? Ideally, the data would be prepared originally in computerized or digital format, and then made widely available to military historians and operations analysts in both computerized and printed tabular form. The printed tables should be prepared directly from the computerized format.

c. **Proposed Data Layout.** The following is a proposed data layout for any future data base of battles. It has obviously been inspired by, and attempts to improve upon, the layout used in existing data bases.

(1) **Identification.** Battle name, dates, and times during which the battle was joined, location(s), order of battle, *etc.*

(2) **General Situation.** Intelligence, objectives, plans, *etc.*

(3) **Particular Situation.** Weather, terrain, deployment, strengths, equipment strengths, supplies available, *etc.*

(4) **Conduct of the Battle.** Maneuvers, intelligence, combat and operational losses, consumption rates, *etc.*

(5) **Outcome.** Status of forces, winner, accomplishments on each side. Narrative summary. Strength and attrition histories. Impact on General Situation.

(6) **Analysis.** Comparison with analogous situations or trends. Lessons learned.

d. The following identifies some particular things to favor or to avoid in data base design and construction.

(1) Avoid the use of ratios. For example, instead of giving the advance rate in kilometers per day, give each advance distance in kilometers, together with the date-time groups at which it began and ended. This provides more information than is present in the ratio. Modern computing machinery can easily determine the ratio as needed, but it is impossible to use the ratio alone to retrieve its component values.

(2) Favor numerical over alphabetical or character data. Although names of things (such as locations, commanders, forces, campaigns, battles, wars, *etc.*) should be recorded in alphabetical or character form, it is possible and often desirable to associate them with numerical codes. Whether or not these inherently alphabetical or character items are married to some numerical coding scheme, it is best to have all other entries numerical, with an explanatory key. In general, mixing numerical and alphabetical or character values is to be avoided, for it complicates the subsequent use of the data base.

(3) Dates and times should be expressed as \pm YYYYMMDDTTTT, where positive values are used for anno Domini (AD) dates and negative values for dates before Christ (BC). Here YYYY is the year, MM is the number of the month of the year, DD is the day of the month, and TTTT is the local time of day in hours and minutes expressed in the military or European 24-hour clock system. The starting and ending date-time groups for the battle as a whole, as well as each of its identifiable phases, must be recorded.

(4) Never use dashes. Instead of using dashes, indicate explicitly what is meant. The reason for this is that dashes are ambiguous. They could mean that the value is either zero; unknown; not applicable (*e.g.*, the indicated entry would be a total of ratios, which often has no interpretable meaning); known and applicable, but not reported for some reason (*e.g.*, it is classified, withheld pending confirmation, *etc.*); not zero, but too small (or too large) to fit the space allotted for it; believed to be known and obtainable, but not currently available; not known and believed to be unobtainable; not known to the same level of accuracy or reliability as other values reported in the table; supposed to be "obvious" (*e.g.*, it may record the date-time group at which a battle ended, with the understanding that dashes repeat the starting date-time group); known and obtainable, but would have to be taken from a different source than the one used for the data given elsewhere in the table; peculiar (*e.g.*, it is from a much earlier or later event than those given elsewhere in the table); the same as in the row immediately above (or to the left of) the dashed entry; and so on and so forth.

(5) Never use blanks. They, too, are ambiguous and can mean almost as many different things as dashes. Again, instead of using blanks, indicate explicitly what is meant.

(6) Particular care should be taken to avoid any use of the letters O ("oh") and I ("ell"), as they are too easily confused with the numbers 0 (zero) and 1 (one), respectively. Reserve the use of 0 (zero) to mean that the value is known, and in fact is known to be zero.

(7) Each coded entry must have a corresponding explanatory key, defining the item encoded and providing the key for deciphering the encoding system.

(8) Special codes need to be used to indicate missing values for each item, and their meaning must be explained. Never use 0 (zero), a blank, or dashes to denote a missing item.

(9) The units of measure (personnel, kilometers, tons per day, barrels per hour, *etc.*) must be specified for each entry.

(10) Limits on the allowable data base entry values need to be explicitly defined. (For example, distances advanced by the attacker on a given day may have an allowable value between -999 and +999 km, and represent the average distance between the original and final lines of contact, expressed in tenths of a kilometer, with positive and negative values corresponding to advances and retrograde movements by the attacker.) This facilitates preparing a computer program to verify that each entry has one of its allowable values. For data bases prepared by contractors, the contract should call for this verification program and its results to be approved by the government as part of the acceptance procedures. Also, the contract deliverables should include this verification program and its results, as well as the computerized and tabular data bases.

(11) At least three different complete keys to interpretation need to be provided: (i) for the computerized data base, (ii) for the printed version of the data base, and (iii) for any computer programs that manipulate the computerized data. Preparing each of these keys requires meticulous attention to detail, precision of statement, and care in maintaining consistency. However, the nature of the keys is different. The computerized data base needs a key to all of the files used, to the file structure (*i.e.*, explaining what records are in the file), and to the type and format of data in each field of the records. For example, the key to the computerized data might describe two files, one of which is a file of battle names and starting dates, while the second is a file of the attacker's initial personnel strengths. The records in each file can be listed in order of some more or less arbitrary battle sequence number. The first file could have one record for each battle sequence number, with each record having two fields. The first field might give the battle name in the form "Monte Lungo North Slope," and be read as an alphabetical string 50 characters long. The second field could give the battle's starting date in the form \pm YYYYMMDD and be read as a signed 8-digit decimal number. The second file could consist of records having a single field and be read as a 7-digit integer number.

The printed version of the data base needs a key by the line numbers on the output printout that names the items displayed on each line. For example, the printed version might read:

Monte Lungo North Slope	22 August 1943
17000	

with the explanation that this is the battle of Monte Lungo North Slope, that started on 22 August 1943, and in which the attacker force initially numbered 17,000 men. Exact, unambiguous, precise definitions need to be provided for each of the variables used. These definitions also need to make a clear distinction between data and judgments.

The variables used in the computer programs need to be clearly defined. The names used for the variables themselves should have mnemonic value (i.e., use KmAdv for distance advanced in kilometers, rather than some nonsense code as BGDK). These mnemonic names must then be used consistently throughout.

(12) Favor continuously variable values to categories. For example, instead of giving force sizes in three broad categories as: (i) small (0 to 10,000), (ii) medium (10,001 to 100,000), or (iii) large (over 100,000), give the force size. With modern computing machinery, the categories can be recreated at will as needed, but the individual values cannot be retrieved from the categories alone. If the individual values are uncertain, then give (separately) an estimate of their uncertainty.

(13) In some cases, categorical variables cannot be avoided. If categories rather than continuous variables must be used, then they must be exhaustive and mutually exclusive. Also, wherever possible, rank categories in some logical or natural order, e.g., by increasing severity or intensity (for example, instead of categorizing seasons as winter, summer, spring, fall, use the order spring, summer, fall, winter).

Also, the natural tendency to create new "mixed" categories must be curbed. For example, after having created defense posture categories such as hasty defense (code HD) and prepared defense (code PD), resist the lazy tendency to enter some defensive posture as HD/PD to indicate that it is some sort of mixed or combined defense posture. Such mixed categories are impermissible. Once categories have been defined (e.g., HD and PD), they must not be mixed. If additional categories must be introduced, then define them explicitly. Also, make sure that the resulting totality of categories is exhaustive, mutually exclusive, and (preferably) monotonic. Redefine and recode the previous categories, if necessary, to achieve these desiderata.

(14) Favor objective measurements over subjective judgments wherever possible. For example, instead of rating morale as good or bad, provide the number of defectors, AWOLs, incidence of mutinous incidents, sabotage, pilfering, and so forth. Where subjective values are unavoidable, they must be clearly identified as such.

(15) Use standard terminology and categorization systems wherever possible. Standardized classification systems have been developed for weather, climate, and terrain (among others). Where such standard systems are available, they should be used in preference to some *ad hoc* or idiosyncratic classification scheme.

2-5. OBSERVATIONS ON ACCURACY REQUIREMENTS. Here we offer some observations on the accuracy of historical data that is needed in order to support subsequent analyses. Errors of at least 5 to 30 percent can be expected in historical combat data bases. These more or less random errors, as well as various possible biases, demand careful consideration in any analysis done on these data bases.

Unfortunately, high accuracy has been found unobtainable in practice. Instead, we must use the data as we find it, warts and all. This is not a new observation. Beebe, Gilbert W., and De Bakey, Michael E., *Battle Casualties: Incidence, Mortality, and Logistic Considerations*, Charles C. Thomas, 1952 (p xi) note that "Military medical statistics which are the product of war are inevitably crude and inexact." McQuie observes that "Despite the voluminous records on file, data available on WWII and Korea are inadequate, incomplete, and probably biased," and that "Most military summaries of the past have been narrative in form with a general tendency to emphasize the good and play down the bad. Unit histories and after-action reports tend to be less than objective. This became especially clear during the data search when an effort was made to reconstruct a particular battle by reviewing the after-action reports starting at the battalion level and then attempting to correlate the data from the three battalions with regimental reports. The three regimental reports in turn bore little resemblance to the division account." For example, *Medical Statistics in World War II*, Office of The Surgeon General, Department of the Army, 1975, notes the following discrepancies for World War II:

<u>Type Casualty</u>	<u>The Adjutant General's Report</u>	<u>The Surgeon General's Report</u>
WIA	592,170	723,560
Total Deaths (KIA + DOW)	216,005	213,030
KIA	189,696	192,220
DOW	26,309	20,810
Other Battle Deaths [†]	18,869	16,793

[†] MIA later declared dead, POW who died in enemy hands, etc.

According to Williams, W. H., "How Bad Can 'Good' Data Really Be?", *American Statistician*, May 1978, pp 61-65, typical error rates of 5 percent to 30 percent afflict even our best data bases. We can expect data bases on historical combat actions to be no more accurate than that—if we are lucky! As noted in Small, Melvin and Singer, J. David, *Resort to Arms: International and Civil Wars, 1816-1980*, Sage Publications, Beverly Hills, CA, 1982, pp 73-74, "Despite these multiple cross-checks and a large dose of skepticism at every turn, we must reemphasize the fact that our battle death figures are only estimates. It is worth bearing in mind the possible sources of erroneous data. First, not all armed forces have been consistent in differentiating among dead, captured, missing, wounded, and deserting. ... Second, there is the simple matter of accurate estimates, compounded by the fact that the size of a force may not be known with any accuracy even by its commander. Third, there are the tactical reasons for exaggerating the enemy's losses and minimizing one's own. Finally, the archivists and historians who

eventually sift through the reports and provide our basic sources of data may well suffer not only from a lack of statistical sophistication but even occasionally from personal and national biases of their own."

Moreover, the accuracy of the data does not necessarily increase with battle date. In particular, eye witnesses are notoriously unreliable, as was noted by Thucydides in his *History of the Peloponnesian War* (translated by Rex Warner, Penguin Classics 1954 and 1972, p 48), who warned that: "And with regard to my factual reporting of the events of the war I have made it a principle not to write down the first story that came my way, and not even to be guided by my own general impressions; either I was present myself at the events which I have described or else I heard of them from eye-witnesses whose reports I have checked with as much thoroughness as possible. Not that even so the truth was easy to discover: different eye-witnesses give different accounts of the same events, speaking out of partiality for one side or the other or else from imperfect memories." This thought is also echoed by Walter Lord in his book, *Incredible Victory*: "The excitement and confusion of battle always breeds conflicts on numbers, distances, hits and misses, the time things happened." Clausewitz, in his famous monograph *On War*, remarks that "Casualty reports on either side are never accurate, seldom truthful, and in most cases deliberately falsified." This is certainly true while a war is in progress, but conscientious historical research often manages to correct the worst of these distortions. As remarked by Melvin Small and J. David Singer in *Resort to Arms: International and Civil Wars, 1816-1980*, Sage Publications, Beverly Hills, CA, 1982, pp 73-74, "Ironically enough ... the post-1945 period gave us more difficulty [*i.e.*, the authors have less confidence in their battle death data] than the earlier period."

Uhorchak, John M., *et al*, "Final Report: Casualty Data Assessment Team Operation DESERT STORM," Letterman Army Institute of Research, Division of Military Trauma Research, Presidio of San Francisco, California 94129, Institute Report No. 469, January 1992, 115 pp, UNCLASSIFIED, complains that the "medical records generally lack pertinent information" such as blood pressure, heart rate, and so forth, and that the category "multiple fragment wounds" (MFW) was used "to describe any wound from a simple 1 to 2 cm fragmentation laceration to a severe traumatic disruption of the popliteal artery."

The difficulties of using primary data are amply illustrated by the following selected quotes from "Average Casualty Rates for War Games, Based on Historical Data," Historical Evaluation and Research Organization (HERO), 15 February 1967, UNCLASSIFIED, available from DTIC as AD-890 546L, AD-890 547L, and AD-890 548L. (p vi) "Where records were inadequate or ambiguous, available figures have been expanded or modified on the basis of professional military and historical judgment. (This was particularly necessary for German data, since all of the most relevant German records have been returned to West Germany without having been microfilmed.)" (p 2) "Enemy records now available in the United States rarely include data for units below division level, and those for US combat elements available for this study were limited in quality and level of resolution by the purpose for which

they were originally compiled. ... Some information is available on the intensity and duration of combat on certain days, but it is no more possible to ascertain the distribution of casualties by category or time than it is valid to assume that they were spread evenly over the duration of the combat." (p 4-5) "No Japanese, North Korean, or Chinese Communist records are available. Consequently, figures on casualties and strengths of those forces were procured entirely from the reports of the opposing US units, and must be viewed with considerable caution. ... In only a few instances were daily reports of German strengths and losses found, and these covered isolated periods of a few days. Much of the material was in the form of monthly reports at the corps or army level. In some cases information pertaining to the same period was found in different forms, although frequently conflicting, necessitating evaluation, and application of professional judgment." (p 6) "In cases where the only German casualty figures available were those accumulated for a stated period, usually ten days or one month ... a daily breakdown of casualties was estimated, based upon knowledge of the situation existing, the nature of the combat in which the units were engaged throughout the period, the intensity of the combat indicated by casualty figures for US forces, knowledge of the course of operations, and experience with similar forces in similar situations." (p 7) "Japanese casualty figures on Okinawa, derived solely from US sources, included only killed, broken down into several categories, including estimated dead as well as counted dead and estimated numbers sealed in caves ... the accuracy of these daily figures is impossible to validate. There are no figures at all on Japanese wounded ... it was found that doubling the number of counted dead, while ignoring other estimated categories, gave the most plausible total for dead and wounded." (p 7) "For Korea, ... we have accepted the Far East Command (FEC) figures ... These evidently do not include estimates of wounded. We have assumed they do." (p A-4) "While no detailed casualty reports of 16th Panzer Division for the entire period were found, total casualties of 1,300 for the division for the Salerno battle were reported by XIV Panzer Corps to Tenth Army. ... These casualties have been somewhat arbitrarily broken down according to the known postures of the division at various times during the period, using professional judgment on the basis of the situation, the known circumstances, and the casualties of the 45th Division during the same periods. (p A-6) "Microfilmed records of XIV Panzer Corps contained several sets of strength and casualty figures for 26th Panzer Division, all of them unfortunately at some variance with each other." (p A-12) "Three separate statements of 45th Infantry Division strength and casualties were found, each of which was in considerable disagreement with the others. The nonavailability of detailed German strength and casualty figures was particularly frustrating in attempting to analyze results of 45th Division operations at Anzio. No strength figures for the German divisions engaged in whole or in part against the division were discovered for February. ... Thus an approximation of each division's strength was arrived at by multiplying the ascertainable infantry strength in each instance by a factor of 5. ... A comparison of the results of this process with secondary historical works ... suggests that the overall German strength was probably somewhat higher than that produced by this method. ... For this reason, in each of the Anzio engagements considered in

detail, an arbitrary 25 percent has been added to the strengths of the German units positively identified, and 25 percent has also been added to their casualty totals." (p A-13) "Casualties of the 45th Infantry Division were reported ... the most recently dated document was used, although there were indications that, in some cases, casualty figures reflected accumulations of casualties of previous days which were reported late. In these cases, clearly erroneous totals have been redistributed among prior days on the basis of known intensity of combat." (p D-6) "The considerations and procedures for development and analysis of north Korean and Chinese Communist strengths and casualties are similar to those for Okinawa, but with some differences. In the first place, opposing overall force strengths and structures are not so well known as was the case on Okinawa. Similarly, the casualties are based mainly on estimates."

This discussion clearly demonstrates that the question "What accuracy is needed to support subsequent analyses?" is improperly posed. Ideally, all values should be reported to better than 5 percent, but this is seldom attainable. For the most part, despite the optimistic views of some data base preparers, the best attainable accuracy is on the order of 5 to 30 percent for strengths and losses. Often the accuracy is much less. Analyses of historical battles and engagements simply must recognize that these data are the most accurate that can be provided, and deal with them as best they can. On the other hand, there is no substitute for continually correcting and improving the accuracy and reliability of the data base.

To a limited extent, a sufficiently large *quantity* of data can partially, but not completely, offset its poor average *quality*, or accuracy. The main proviso for this is that the available data be a *representative* sample, *i.e.*, that they be unbiased. When this is the case, appropriate statistical methods can take maximum advantage of the additional information provided by a large quantity of not very accurate data.

The problem of systematic errors, or biases, that make the data nonrepresentative is a thorny one. Such biases may arise in various subtle ways, such as in selecting the battles to include/exclude and in tabulating the data on individual battles. Since even slight biases have been known to vitiate statistical results, the issue is too important to ignore, although its importance has often been underrated. The primary rule is that quality comes before quantity in compiling data. Cochran, William G., *et al*, *Planning and Analysis of Observational Studies*, John Wiley & Sons, 1984, offers the worthwhile suggestion that "... the investigator who has a nonrandom [*i.e.*, nonrepresentative] sample must envision the kind of population from which the sample might be regarded as drawn at random. It is helpful to give a name—the *sampled* population to this kind of population and to describe the ways in which it differs from the *target* population about which we would like to draw conclusions." Biases may arise in the following ways:

a. Preferential selection of stunning victories, egregious follies, incredibly stubborn defenses, fanatical attacks, amazing upsets, and similar noteworthy—and by that very fact extraordinary—battles. This is a kind of “Guinness Book of Records” approach. All of military history is inclined to do this, so deliberately leaning against the flow is necessary to maintain a proper equilibrium.

b. Preferentially selecting a series of battles where one side was on a streak of victories. Examples are the Allies near the end of WWII, the earlier Napoleonic era, the Union near the end of the Civil War, the Battles of Frederick the Great, and the Israelis in recent mid-Eastern actions.

c. Tending toward awarding victory to the “Great Captain of Revered Reputation” rather than on the basis of the actual outcome, or of preferentially selecting battles that highlight the talent of genius at the expense of those showing the hero with feet of clay.

d. Unfairly counting POW/MIA as casualties, thus exaggerating the losses to the side that was defeated by a great captain, while on his side only those actually KIA are listed as lost—even though his WIA and DOW may be pretty high.

e. Most of the data bases available in the Western world clearly are not representative of all battles fought. They do not contain data on many of the smaller, lesser-known or little-remarked battles. This is particularly true of battles that took place in Asia, Africa, the Middle East, South America, or the Pacific Ocean islands. Furthermore, they are data bases of land battles and do not include air or sea battles. Even for land battles, they often exclude sieges. They offer scant coverage of limited, guerrilla, or insurgent warfare. They rarely include battles from the Russian Revolution, Russian Civil War, Korean, Malaysian, Algerian, Arab-Israeli, Soviet-Afghanistan, Indo-Pakistan, Iran-Iraq, or Vietnamese Wars. They are predominantly data bases of division- or corps-sized forces engaged in short (hours to days) intense clashes on the European or North American continents during the 19th and 20th centuries, rather than data bases of wars or campaigns conducted at the operational level.

f. Most of the data are on battles. Some data are available on wars, but little or no data are available on campaigns. This may bias the analysis in unforeseeable ways.

g. Often battle durations are given in days. This is much too coarse a time scale. Battles which last for less than a day, or which stretch out over just 2 or 3 days, have durations that are badly misrepresented by this coarse a time scale. The coarseness of the time scale makes all calculated rates essentially meaningless for short battles. For example, take two battles in which the personnel strengths, casualties, and distances advanced were the same, but one battle lasted 1.5 hours while the second lasted 15 hours. If durations are given in days, both would be listed in the tables as having the same percent casualties per day and the same rate of advance. Yet the first battle actually had rates 10 times higher than those of the second battle—a fact that is concealed by the tabulated data.

h. Battles should be subdivided on a consistent basis. For example, one data base divided the Okinawa Campaign into 28 separate battles, and Iwo Jima into 3. Yet the same data base treated the Defense of Moscow as a single battle, although it lasted 65 days, spanned a 700-km front, and involved well over a million troops on each side (almost 2.5 million altogether). This discrepancy grotesquely overrepresented operations by the US versus Japan in the Pacific at the expense of operations by the Soviets versus Nazis in Europe. The amount of bias introduced by this over-representation is difficult to assess.

i. It is impossible to maintain consistency in a data base without exact operational definitions of terms. For example, total battle casualties are supposed to include KIA, WIA, MIA, and POW. But in some cases, the data available do not permit this. Accordingly, an indication should be provided to distinguish which battles list battle casualties according to the above definition and which use some other basis (such as reported "losses," which include who-knows-what). Such indications need to be provided for both sides, so that battles having casualties listed one way on one side (*e.g.*, KIA) and a different way for the other side (*e.g.*, total battle casualties) can be recognized as such.

j. Personnel strengths should be provided on the same basis for both sides. In one data base, personnel strengths were labeled as "Total Engaged," although in some cases (*e.g.*, Ypres III and the Somme) the battle casualties exceeded the "Total Engaged." This came about because the "Total Engaged" values were actually average daily strengths. Unfortunately, this makes it impossible to recover either the actual initial strengths, or the total reinforcements/replacements during the battle, or the final strengths. That method of reporting the data was a bad mistake. The average daily strength can easily be estimated from the initial strengths, the casualties, and the final strengths—which are the values that should have been reported.

2-6. DEALING WITH DEGREES OF CONFIDENCE. Here we offer some observations on how to deal with varying degrees of confidence in the data. Although it is an unfamiliar area to most military historians, degrees of confidence in the data base values should be developed and reported.

a. First, degrees of confidence in the data definitely should be estimated by those assembling it. Otherwise, users will have no guide as to whether data for a given battle is considered "firm, accurate, and trustworthy," versus "soft, approximate, and suspect." By presenting estimated degrees of confidence in the form of a "best estimate," together with its associated upper and lower "confidence limits" in the fashion described in the statistical literature on tolerance intervals and personal probability, users will be able to distinguish rough estimates from more accurate values.

b. Such uncertainty estimates are proper and should be provided for objective and numerically valued quantities. On the other hand, such uncertainty estimates are probably not worthwhile for frankly subjective values (such as the degree to which one side or the other had superior morale). Also, uncertainty estimates probably cannot be satisfactorily expressed for objective quantities that are

categorical in nature (such as the protective posture of defending troops). In addition, subjective assessments of uncertainty cannot be expected to encompass the likelihood of typographical and arithmetic mistakes, or other oversights and mental slips—these are best detected by internal consistency checks.

c. Experience with assessing uncertainties that have demonstrated that few military historians are well-equipped to deal with this conceptually because they lack the skills needed to make it a routine operation. Most of them resist the basic concept. Often, they are overconfident in the accuracy of their reported data. Nevertheless, it is essential that such confidence estimates be provided. Perhaps a comparison of the values published by a given historian or historical analysis group at different times and venues would shed light on the true levels of uncertainty in the data. Also, error estimates may be developed by considering the various possible error sources or components, such as systematic errors (biases), data transcription and handling errors, and other random errors.

d. Astronomers are familiar with the concept of a “personal equation,” whereby the observations of individual astronomers are systematically corrected or “adjusted” to agree more closely with those of the average. A similar effect is likely present in military historians, even though as yet no “personal equations” have been developed to correct for their biases. However, it has been demonstrated in the psychological literature that people generally are much more confident in the accuracy of their own estimates than is justified on the basis of objective measurements. Such biases must be recognized and borne in mind when appraising subjectively the accuracy of the values in a data base of battles. Perhaps a comparison of the values published by different historians or historical analysis groups would shed light on the nature of potential biases.

e. The data base may be inadvertently biased. Some of the biasing tendencies that may or may not be present include the following:

- Victory, especially in “close calls,” may tend to be awarded to famous leaders, to the side with the highest force ratio, or to the “traditionally superior” nation (*e.g.*, USA, Germany, Israel).
- History tends to be written by the victors.
- Some of the data may be proprietary and hence not subject to close scrutiny by other qualified historians.
- There may be a “dominant analyst” whose interpretations and views prevail, thus exacerbating the possibility of idiosyncrasies and the “personal equation” affecting the data.
- There is a well-known tendency to be overconfident of one’s own judgment, which may in particular bias downward all estimates of the frequency and size of the errors in the data.

Observe that purely internal consistency checks are not likely to catch systematic biases in the data. Only “external” checks by independent analysts, or the use of other data bases gathered by other investigators or at other times and places, can do that.

2-7. QUALITY CONTROL MEASURES. The values in data bases intended for scientific use should be systematically checked by using computer programs to verify that all of the individual values are legal, and that various combinations of values are mutually consistent.

a. The first thing to check is that the advice given in the previous paragraphs regarding the data that should be included, how it should be presented, and its confidence level has been followed.

b. Detailed internal consistency checks should be made. In particular, a computer program should be written to check that all data entries have values that are permissible for the kinds of information they convey. (For example, that no unit has a negative number of personnel in it, that its casualties do not exceed its initial number of personnel plus any increases by reinforcement or other means, *etc.*). Either the same or a separate program should also be written to check for inconsistencies among different data entries (*e.g.*, that the defense is awarded a victory even though the attacker made a rapid and deep advance, or that the attacker is awarded a victory even though he failed to advance against the defender). Careful use of such programs to check for illegal individual and inconsistent multiple data base values can be a big help in purging a data base of typographical errors, missing values, *etc.* For this approach to be feasible, it is essential that the data base be computerized. In contracting for the development of a data base, it should be specified that the contractor check it for consistency using a computer program, that the consistency-checking program and its results be approved by the government, and that contract deliverables include the consistency-checking program, as well as the computerized data base, so that the government can run it against the data base and determine whether or not it gets the same outputs reported by the contractor.

Careful proofreading of the data base against narrative accounts can also point up questionable data values for further examination. In some cases, graphing the data (for example, when the data include location coordinates) can be useful in flagging values for further inspection.

However useful such consistency checks may be in helping to identify "suspicious" data values, they alone cannot check on the factual validity of the data values. For that, additional work (at least spot checks) by independent historical analysts is indispensable.

c. Make sure that terms are clearly defined, that standard definitions are used whenever possible, and that all definitions have been followed conscientiously.

d. Check to see that all of the battles included are sufficiently similar. (If most of the battles in a data base are of division or corps size, then it is inappropriate to include in it such Churchillian-sized "battles" as the Battle of France.)

e. Make sure that subjectively defined items are clearly identified as such.

f. Verify that there is an audit trail for each data item. For each data entry, is there a way to tell how it was obtained? For example, was it extracted directly from some source, and if so, what source was used? If not extracted directly from some source, how was the value obtained (on what information is it based, and how was that information used to generate the reported value)?

2-8. ADDITIONAL DATA NEEDS. Here we offer some observations on the need for additional data, and on how it might be obtained. Additional data bases are needed to provide attrition histories, detailed loss data, information on campaigns, and include a wider range of time, space, and military situations. Consideration should be given to a larger use of data from field tests, experiments, and training exercises. The military services in general, and the US Army in particular, should lay plans for collecting data on operations in future combat situations.

a. The most critical need at the moment is for several attrition histories. By this we mean data on the exact timing of losses and reinforcements in enough detail to determine the strengths and losses as a function of time for both sides in a number of specific battles. The available data bases of this type are few in number and suffer from a number of various deficiencies. This sort of data is absolutely essential for any rational examination of the laws governing attrition in combat. More and better data is required for progress in this important area.

b. Most of the available data bases give only one type of loss. Often the type used is total battle casualties, defined to be the sum of the KIA, WIA, MIA, and POW. However, many potential applications require data for each of the separate casualty categories included in total battle casualties. Indeed, for some applications, losses need to be broken down by finer categories, such as service branch, military occupational specialty, and so forth. As a minimum, losses should at least be separated into the bloody (KIA and WIA) and nonbloody (MIA, POW, or DNBI) categories, since it is often considered that the bloody losses are the ones most relevant to victory and defeat, while the nonbloody losses are generally considered to be dependent upon the bloody losses.

c. Most of the existing data bases can fairly be characterized as data bases of battles, rather than of campaigns or wars. A few of them do deal with wars, but we are not aware of any that deal specifically with campaigns, i.e., for operations at the grand tactical or operational level. The nearest corresponding items probably are the West Point Atlas of American Wars and some of the Soviet studies of the operational level of warfare, although neither of these is really a data base in which the same items of information are assembled on each of a number of campaigns. Extensive planning and preparation for any such data base will be needed for best results.

d. Most of the data bases available in the Western world clearly are not representative of all battles fought. In particular, they do not contain data on many of the smaller, lesser-known or little-remarked battles, such as those typical of limited, guerrilla, or insurgent warfare. Even for land battles, they often exclude sieges. They are predominantly data bases of division- or corps-sized forces engaged in

short (hours to days) intense clashes on the European or North American continents during the 19th and 20th centuries, rather than data bases of wars or campaigns conducted at the operational level. It would be desirable to expand these data bases to include a wider range of time, space, and military situations. In this, attention should be given to the smaller, lesser-known, and little-remarked battles, so that we can see whether they have the same characteristics as more famous battles. Coverage of air and sea battles would also be welcome. While wars are covered to some extent by existing data bases, they often provide scant information on the nature and activities of the combat forces involved.

e. Even enlarging the existing data base of European and North American battles would be worthwhile. Several analyses that should be performed cannot be done because the number of battles is too limited. This occurs whenever data are to be categorized according to several features. For example, characterizing battles by size, by mix of infantry and other branches, by terrain, by weather, and by battle date soon produces a table with numerous missing entries because of the lack of enough data items. This makes interpretation of the data difficult or impossible.

f. It is often mentioned that the data obtained during field tests, experiments, and training exercises, such as those conducted at the Combat Developments Experimentation Center or the National Training Centers, may be useful in helping to understand the historical data. Although few efforts have been made to do so, it would be interesting to determine whether this suggestion can effectively be put into practice. If nothing else, these experimentation centers may provide a good environment for training military historians or the other independent combat data collectors how to collect detailed data during actual combat operations. In addition, it may be possible to "calibrate" the field experiments to historical data (as suggested by David Rowland of the United Kingdom's Defence Operational Analysis Centre (DOAC)—see Rowland, David, "Assessments of Combat Degradation," RUSI, June 1986.).

g. The Army historically has been ill-prepared to gather scientific data on its operations. Consequently, when the need arises, there are no plans, no cadre, no training programs or manuals, no procedures worked out, no support requirements identified, no good ideas of what equipment or technology can be exploited, no notions of how many people or what kinds of skills are needed, no arrangements for where the needed people are going to come from, no objectives or priorities established for what kinds of information to collect, no concepts of what to do with it once collected, no lessons learned from prior experience, and everything has to be slapped together in a mad rush. Excellent opportunities for debugging this kind of combat data collection system were offered by the Falklands, Granada, and Panama operations, but nobody even thought to use them for that purpose. The penalty for such lack of preparation is underscored by a remark of General O. P. Weyland's regarding World War II to the effect that what happened was not clearly perceived, or if clearly perceived, was not remembered, or if remembered, was not written down, or if written down, was not disseminated, or if disseminated, was not read, or if read, was not understood.

Personnel with experience in monitoring and evaluating the performance of military units (exercise controllers, training exercise evaluators, maneuver controllers, *etc.*, as well as wargamers, operational test and exercise evaluators, and some practical or applied statisticians) should be identified, organized, tasked with fleshing out the above suggestions with practical implementation methods, and assigned to carry out the resulting plan. In the meantime, stocks of necessary materials for use by observers and data collectors should be assembled. These include hand calculators, portable computers, watches, clipboards, walkie-talkies, associated software packages, and spare batteries. Some ways also need to be worked out for gathering, filing, and preserving in good order for later rapid retrieval and analysis whatever electronic information is gathered. What form this should be in (hard copies and permanent magnetic storage such as floppy disks, *etc.*) needs to be decided. Also, requirements for transportation and logistical support for data collectors should be identified and their magnitude and timing planned for.

As an important part of the overall data collection activities, special care should be given to document and to write the history of each data collection effort, with particular attention to the lessons learned from it.

h. We also need to seek out and experiment with better ways of recording, retrieving, and analyzing ill-formatted materials (such as maps and operational overlays, pictures, audio and/or visual recordings, narrative accounts, notes, samples of equipment and ordnance materiel, blueprints, and similar materials).

2-9. SUITABILITY OF PERSONNEL SKILLS AND CHARACTERISTICS. Here we offer some observations on the kinds of personnel skills and characteristics best suited to the construction and maintenance of historical data bases on land combat operations. The planning, construction, and analysis of an extensive historical combat data base is a difficult undertaking, demanding the coordination of a wide variety of skills in a multidisciplinary effort. Therefore, such data bases as exist should be considered as valuable assets to be carefully preserved, maintained, and made widely available for study and analysis.

a. Who should be involved? It is important to involve in data base planning and creation military operations analysts and statisticians, as well as military historians and archivists, and computerized data base experts. The kinds of expertise developed to deal with similar data collection and analysis problems in other fields should also be exploited (examples are economics, sociology, anthropology, actuarial statistics, and psychology as well as the hard sciences of physics, chemistry, and engineering). Medical, personnel, logistics, and other specialists should also be involved in areas where their expertise is applicable. Epistemologists also have important advice to give on this subject.

b. The planning, construction, and analysis of an extensive data base of battles, campaigns, or wars is a difficult undertaking, demanding a wide variety of skills, and requiring a multidisciplinary approach. Obviously, military historians need to be involved. However, other skills and professional disciplines are also essential to good results. The whole should be under the direction of an individual with some familiarity with each of the essential disciplines involved, which probably means an experienced military operations analyst. He, in turn, needs to call upon military historians, professional soldiers (in all of the usual staff categories), statisticians, sociologists, psychologists, epistemologists, and other for skills in specific supporting areas. Advisors familiar with practical data base development in the political sciences, economics, or the hard sciences can also contribute valuable advice and suggestions.

c. Since such data bases are so costly in terms of money, time, and talent, all such data bases should be considered as valuable assets. Special attention should be given to preserve them, maintain them, and make them as widely available for study and analysis as possible.

2-10. SUMMARY OF MAIN POINTS. In this chapter we have focused on historical combat data bases suitable for use in quantitative scientific analyses of military operations. The principal lesson learned is that the preparation of battle and engagement data bases adequate for use in quantitative analyses is much harder to do well than seems to be generally appreciated. There are manifold conceptual, technical, and mechanical difficulties to be overcome. As a result, it is far more costly and time-consuming than might be expected. Accordingly, careful planning and preparation are needed in order to maximize the results from a given level of resources. Since results of considerable value in military operations analysis and wargames can be obtained from a clever analysis of a carefully prepared data base, more effort should be devoted to pursue this approach.

In this regard, we express our opinion that the CDB90-1991 computerized data base is far superior to other available data bases on battles and engagements in its availability, length of time span covered, number of battles included, detailed coverage of individual battles, and accuracy. It nevertheless has many important flaws and shortcomings, and additional effort is needed to address them.

Summary responses to the subordinate issues are given below:

a. Data bases may draw upon primary, secondary, or tertiary sources. Quantitative analysis uses historical data in ways that can be distinguished from most past uses.

b. Lacking an adequate theoretical basis, the selection of data to record is often *ad hoc*. Some guidance can be provided by erecting specific hypotheses to be tested, and then assembling data suited to such tests. Careful plans for data collection will get the best results from a given level of resources.

c. Systematically organized data bases, with the same data items provided for each of a large number of cases, with all data developed on a common basis for comparability and compatibility, are needed. These data can easily be manipulated for analysis purposes if provided in computerized format. Hard copy versions could be prepared directly from the computerized versions.

d. Errors of at least 5 to 30 percent can be expected in historical data bases. These more or less random errors, as well as various possible biases, should be carefully considered in any analysis done on these data bases.

e. Although it is an unfamiliar area to most military historians, degrees of confidence in the data base values should be developed and reported.

f. The data base values should be systematically checked by using computer programs to verify that all of the individual values are legal, and that various combinations of values are mutually consistent.

g. Additional data bases are needed to provide attrition histories, detailed loss data, information on campaigns, and include a wider range of time, space, and military situations. Consideration should be given to a larger use of data from field tests, experiments, and training exercises. The military services in general, and the US Army in particular, should lay plans for collecting data on operations in future combat situations.

h. The planning, construction, and analysis of an extensive historical combat data base is a difficult undertaking, demanding the coordination of a wide variety of skills in a multidisciplinary effort. Therefore, such data bases as exist should be considered as valuable assets to be carefully preserved, maintained, and made widely available for study and analysis.

(THIS PAGE INTENTIONALLY LEFT BLANK)

APPENDIX A

LIST OF DATA BASES

A-1. INTRODUCTION

a. This appendix lists what appear to be the main data bases on battles and wars available on diskettes usable with personal computers. The chief criteria for inclusion in this list, and hence in Appendix B's catalog of data bases, are as follows (roughly in order of importance):

(1) In the public domain, so that copies can be made available to governmental agencies and others without restriction and for (at worst) a nominal cost. However, some important proprietary data bases are also mentioned for the sake of completeness.

(2) In data base form (tabulations rather than narratives).

(3) Contain information on military operations in and/or outcomes of battles or wars.

(4) Available on diskettes that can be used with personal computers.

(5) Useful to many military operations analysts; developers, users, assessors, and validators of the inputs and/or outputs of war games and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

(6) Difficult or inconvenient for individuals and separate study teams to generate or recreate, but which would be used frequently if readily available through DTIC or NTIS.

b. The data bases included in Appendix B's catalog are listed below in alphabetical order by their short title. These short titles are used for ease of reference. For example, ACSDB-1990 refers to the Ardennes Combat Simulation Data Base, specifically to the version that was current or created in the year 1990. The long or full title is given immediately following the short title.

c. There are other data bases that deserve to be mentioned in this paper, even if for some reason they do not satisfy all of the criteria for inclusion in Appendix B's catalog of data bases. For example, a data base might be proprietary, or otherwise require special handling; or it may not be relevant to personnel attrition rates in land combat operations (though providing other information of interest to military operations analysts); or it may not have been reducible to electronic form within the level of effort available to the PAR studies.

Some examples of proprietary data bases are those maintained and provided to researchers by the Inter-University Consortium for Political and Social Research (ICPSR). This organization's goals and procedures are described briefly in Appendix C, together with some of its data bases that seem most likely to be of interest to military operations analysts.

Several other data bases or potential sources of information are described in Appendix D.

A-2. LIST OF DATA BASES DESCRIBED IN APPENDIX B

ACSDDB-1990, "The Ardennes Campaign Simulation Data Base," Data Memory Systems Incorporated Report, unnumbered, 18 December 1989, UNCLASSIFIED, work done under Contract No. MDA903-87-C-0787. Report available from DTIC or NTIS as AD-A240 088. Diskette available from DTIC or NTIS as AD-M000 074 (33 diskettes, UNCLASSIFIED, February 1990).

ADE-1992, Helmbold, Robert L., "Ardennes Division Engagement Data," US Army Concepts Analysis Agency unpublished material, 1992, UNCLASSIFIED. Data base on diskette provided to DTIC concurrently with this document.

AHART-1993, "Analysis of Historical Artillery Expenditures (AHART) Study-CY87," US Army Concepts Analysis Agency Technical Paper CAA-TP-87-6, June 1987, UNCLASSIFIED. Report available from DTIC or NTIS as AD-F860 140. Data base on diskette provided to DTIC concurrently with this document.

BENCHMARK-1988, "Data on Historical Battles 1937-1982 (Benchmarks)," US Army Concepts Analysis Agency Research Paper CAA-RP-87-2, July 1988, 60 pp, UNCLASSIFIED. Report available from DTIC as AD-A200 036. Diskette available from DTIC as AD-M000 000 or AD-M000 092 (1 diskette, UNCLASSIFIED).

BERNDT-1993, adapted from Berndt, Otto (Captain in the Austrian General Staff), *Die Zahl im Kriege: Statistische Daten aus der Neueren Kriegsgeschichte in Graphischer Darstellung*, [Number in War: Statistical Data from Modern Military History in Graphical Form], G. Freytag & Berndt, Vienna, 1897, 169 pp, UNCLASSIFIED, available from US Army Command and General Staff College Library (355.09 B524z). Data base on diskette provided to DTIC concurrently with this document.

BOB18-1993, adapted from Helmbold, Robert L., "Air Battles and Ground Battles—A Common Pattern?", RAND Corporation Paper P-4548, 1971, UNCLASSIFIED, available from DTIC as AD-718 975. Data base on diskette provided to DTIC concurrently with this document.

BODASHIP-1993, a collection of naval battle data taken from Bodart, Gaston, *Militar-Historisches Kriegs-Lexicon (1618-1905)*, C. W. Stern, Vienna and Leipzig, 1908, 956 pp, UNCLASSIFIED, available from libraries. Data base on diskette provided to DTIC concurrently with this document.

BWSH-1993, a collection of land battle data taken from Bodart, Gaston, *Militär-Historisches Kriegs-Lexicon (1618-1905)*, C. W. Stern, Vienna and Leipzig, 1908, 956 pp, UNCLASSIFIED, available from libraries. Data base on diskette provided to DTIC concurrently with this document.

CDB90-1991, "CAA's data base of Battles—Version 1990," US Army Concepts Analysis Agency, unnumbered, UNCLASSIFIED, diskette available from DTIC as AD-M000 121 (two diskettes, UNCLASSIFIED, 30 April 1991).

CEC-1933, adapted from Brinkerhoff, John R.; Dupuy, Trevor N.; Johnson, C. C.; and Bader, Brian R., "Casualty Estimates for Contingencies (CEC)," Data Memory Systems, Inc., Report No. CEC-0002AM, 15 November 1985, UNCLASSIFIED, in two volumes (41 pp and 211 pp). Report available from DTIC as AD-A183 255 (Vol 1) and AD-A183 256 (Vol 2). Data base on diskette provided to DTIC concurrently with this document.

COMP-1983, Compton, Jonathon, "Personnel Attrition Data," US Army TRADOC Systems Analysis Activity, White Sands Missile Range, New Mexico, Technical Report TRASANA, TR-29-83, ACN 67513, November 1983, 229 pp, UNCLASSIFIED-Limited, available from DTIC (AD-B085 206L). Data base on diskette provided to DTIC concurrently with this document.

CRETE-1993, adapted from Engel, Joseph H., "Combat Effectiveness of Allied and German Troops in the World War II Battle of Crete," Operations Evaluation Group Interim Research Memorandum IRM-35, 8 February 1963, UNCLASSIFIED, available from Center for Naval Analyses. Data base on diskette provided to DTIC concurrently with this document.

CREWCAS-1991, "Injuries to Tank Crews," Centre for Operational Research and Defence Analysis (CORDA Ltd), 233 High Holborn, London WC1V 7DJ, England. Report Number YR 4373. Work sponsored by the US Army Model Improvement and Study Management Agency (MISMA), 1725 Jefferson Davis Highway, Suite 808, Crystal Square 2, Arlington, VA 22202. Report available from MISMA in the form of camera-ready copy. Diskette available from DTIC as AD-A265 499 (five diskettes, 3-1/2 inch, 730 KB, UNCLASSIFIED, March 1991).

DODGE-1993, Dodge, Theodore Ayrault, *Napoleon: A History of the Art of War from Lutzen to Waterloo, With a Detailed Account of the Napoleonic Wars*, (Vol 4 of four volumes), AMS Press, Inc., New York, NY 10003, 1970, reprinted from the edition of 1904-1907, UNCLASSIFIED, available from libraries. Data base on diskette provided to DTIC concurrently with this document.

ICPSR-1984, "Wages of War, 1816-1980: Augmented With Disputes and Civil War Data," ICPSR data base number 9044. Inter-University Consortium for Political and Social Research (ICPSR), P.O. Box 1248, Ann Arbor, Michigan 48106, UNCLASSIFIED-Limited, US Department of Defense components can obtain a copy from DTIC (AD-M200 224, distribution authorized to DOD components only; proprietary data; January 1994.) Others must obtain a copy from ICPSR.

INCHON-1993, Busse, James J. "An Attempt to Verify Lanchester's Equations," Paper 9-1 (pp 587-597) in Volume 2 of *Developments in Operations Research*, Benjamin Avi-Itzhak (Executive Editor), Gordon and Breach Science Publishers, New York, 1971 (ISBN 0-677-30840-X), UNCLASSIFIED, available from libraries. Data base on diskette provided to DTIC concurrently with this document.

IWOJIMA-1993, Helmbold, Robert L., "Iwo Jima Spreadsheet," US Army Concepts Analysis Agency, November 1993, UNCLASSIFIED, unpublished personal notes. Data base on diskette provided to DTIC concurrently with this document.

KELLEY-1977, Kelly, C. T., *et al.*, "Quality-Quantity Tradeoffs: A Historical Analysis of Air Combat," The RAND Corporation Report R-1976-AF, April 1977, 45 pp, UNCLASSIFIED. Report available from DTIC as AD-C010 504. Data base on diskette provided to DTIC concurrently with this document.

KRIVO-1993, selected extracts from Krivosheyev, G. F. (general editor), *Losses to the Armed Forces of the USSR in Battles, Combat Operations, and Military Conflicts: Statistical Investigations*, Military Press, Moscow, 1993, (in Russian), UNCLASSIFIED (formerly SECRET), 415 pp, book available from publisher and libraries. Data base on diskette provided to DTIC concurrently with this document.

LIVERMORE-1993, adapted from Livermore, Thomas L., *Numbers and Losses in the Civil War in America, 1861-65*, Kraus Reprint Co., Milwood, NY, 1977 (a reprint of the Indiana University Press edition of 1957, originally published in Boston in 1900), 150 pp, UNCLASSIFIED. Available from publisher and libraries. Data base on diskette provided to DTIC concurrently with this document.

LMI-1989, Kuhn, George W. S., "Ground Forces Battle Casualty Rate Patterns: The Empirical Evidence," Logistics Management Institute Report FP703TR1, September 1989, 270 pp, UNCLASSIFIED. Report available from Logistics Management Institute (LMI), 2000 Corporate Ridge Parkway, Tysons Corner, VA 22102. Data base on diskette provided to DTIC concurrently with this document.

MOTTELAY-1993, Mottelay, Paul F., and T. Campbell-Copeland, *The Soldier in Our Civil War: Columbian Memorial Edition*, Stanley-Bradley Publishing Co., 1893, UNCLASSIFIED, in two volumes (475 pp and 444 pp, respectively). Data base on diskette provided to DTIC concurrently with this document.

ORALFORE-1972, Dupuy, Trevor N., and Hayes, Grace P., "Opposed Rates of Advance of Large Forces in Europe (ORALFORE)," Historical Evaluation and Research Organization (HERO), 28 August 1972, UNCLASSIFIED, prepared by HERO for the US Army Deputy Chief of Staff for Military Operations (DAMO-PLW) under Contract No. DAHC19-72-C-0015, available from DTIC as AD-902 830. Data base on diskette provided to DTIC concurrently with this document.

PARMISC-1993, miscellaneous personnel attrition data assembled from various sources, 1993 UNCLASSIFIED. Data base on diskette provided to DTIC concurrently with this document.

POGOGORO-1933, adapted from Historical Evaluation and Research Organization report "Historical Scenarios of Soviet Breakthrough Efforts in World War II," April 1978, sponsored by the USAF Tactical Fighter Weapons Center, Nellis AFB, Nevada, under Contract No. F49620-77-C-0026. Data base on diskette provided to DTIC concurrently with this document.

REISTERK-1986, selections from Reister, Frank A., *Battle Casualties and Medical Statistics: U.S. Experience in the Korean War*, The Surgeon General, Department of the Army, Washington, DC, 1986, 172 pp, UNCLASSIFIED, book available from libraries. Data base on diskette provided to DTIC concurrently with this document.

RMC-1993, Cockrell, James K.; and Carter, Donn, "Research Study on Predictive War Game Factors," RMC Research Corporation, unnumbered report, March 1974, 251 pp, UNCLASSIFIED, available from the Technical Centre of the Supreme Headquarters, Allied Powers in Europe (SHAPE), prepared under SHAPE Technical Centre Contract C.72-03. Data base on diskette provided to DTIC concurrently with this document.

ROADATA-1990, "A Compilation of Data on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-4. February 1990, UNCLASSIFIED. Report available from DTIC as AD-A220 426. Diskette available from DTIC as AD-M000 115 (2 diskettes, UNCLASSIFIED, February 1990).

SP128-1961, "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group, Staff Paper CORG-SP-128, 1 July 1961, 178 pp, UNCLASSIFIED. Report available from DTIC as AD-480 975. Data base on diskette provided to DTIC concurrently with this document.

SP190-1964, "Historical Data and Lanchester's Theory of Combat: Part II," Combat Operations Research Group Staff Paper CORG-SP-19. August 1964, 132 pp, UNCLASSIFIED. Report available from DTIC as AD-480 109. Data base on diskette provided to DTIC concurrently with this document.

VOEVOD-1993, adapted from Voevodsky, John, "Quantitative Analysis of Nations at War," Peace Research Reviews, Vol 4, 1970, pp 1-63, UNCLASSIFIED, published by Canadian Peace Research Institute, Oakville, Ontario, available from libraries. Data base on diskette provided to DTIC concurrently with this document.

WALKER-1991, "Armored Vehicle Weapon Impact Assessment in Southeast Asia," Air Force Institute of Technology Thesis, AFIT/GOR/ENS/91M-20, 290 pp, UNCLASSIFIED/LIMITED. Document available from DTIC as AD-B156 495. Diskette available from DTIC as "Walker Crew Casualty Data Base," AD-M200 173L.

WESTWALL-1993, adapted from data in Historical Evaluation and Research Organization (HERO) Report Number 129, "Combat History Analysis Study Effort (CHASE) Data Enhancement Study (CDES)," 31 January 1986, in five volumes, UNCLASSIFIED, prepared for CAA under Contract Number MDA903-85-C-0252 and available from DTIC under the following accession numbers: AD-A175-712, AD-A175-713, AD-A175-714, AD-A175-715, and AD-A175-716; and in unnumbered HERO Report "Data Base Error Correction (DBEC)," 23 January 1987, UNCLASSIFIED, prepared for CAA under Purchase Order Number MDA903-86-M-8560 and available from DTIC under accession number AD-A176-750. Data base on diskette provided to DTIC concurrently with this document.

APPENDIX B

CATALOG OF DATA BASES

B-1. INTRODUCTION. This appendix contains our catalog of data bases, arranged in alphabetical order by their short title as explained in Appendix A. Those short titles provide the key to the data bases, and are used consistently throughout the rest of this paper.

B-2. EXPLANATORY REMARKS

- a. See the Glossary for definitions of terms and abbreviations.
- b. Text included in square brackets, [like this] are clarifications or comments interpolated by the compiler of this bibliography and are not part of the original text.
- c. In general, more information is given on data bases that are larger and organized in a more complex fashion. Smaller and more simply organized data bases require little explanation.
- d. We have included a few data bases that, technically, do not fit the criteria listed in paragraph 1-3 of the main body. To give a few examples, we have in mind data bases such as AHART-1993 (does not contain tabulated data on casualties or attrition) and BOB18-1993 (does not deal with land combat forces). These and similar works came to our attention while compiling the list of data bases, and it seemed appropriate to include them. However, we hereby put the reader on notice that we made no systematic attempt to seek out such data bases and that there must be many other such data bases that did not happen to come to our attention.

ACSDB-1990

1. **General Description.** "The Ardennes Campaign Simulation Data Base," Data Memory Systems Incorporated Report, unnumbered, 18 December 1989, UNCLASSIFIED, work done under Contract No. MDA903-87-C-0787. Report available from DTIC or NTIS as AD-A240 088. Diskette available from DTIC or NTIS as AD-M000 074 (33 diskettes, UNCLASSIFIED, February 1990).

2. **Data Base Description.**

a. **General.** The Ardennes Campaign Simulation Data Base (ACSDB) is a data base of the World War II campaign fought in December 1944 and January 1945, popularly known as the "Battle of the Bulge," and is provided in both hard copy and computerized form. The hard copy documentation contains bibliographic, definitional, and other information; a user's guide for the ACSDB; a narrative on the single shot probability of kill (SSPK) data researched for the ACSDB; and a collection of photocopies of all records used for generation of the ACSDB totaling some 10 cubic feet of printed material. The computerized version of the ACSDB is about 39 megabytes in size and uses the dBASE IVTM data base management system to facilitate storage and manipulation. American, German, and British combat units that were significantly involved are included.

b. **Information Provided.** The data include daily information on unit location, activities, order of battle, personnel, equipment, and logistics at the army, corps, division, and brigade level for the period 16 December 1944 through 16 January 1945, plus data on tactical air operations, tables of organization and equipment (TOE) of units, and equipment used by opposing forces in the Ardennes Campaign. The casualty categories used are KIA, WIA, and CMIA.

c. **Situational Descriptors Included.** None, other than the name and year of the engagement.

d. **Data Sources Used.** The data were obtained from both primary and secondary sources.

(1) The primary records used for daily information on front line locations, boundaries between front line units, and other unit location data included US Army daily SITREPs (situation reports) and British Army War Diaries, and OKW briefing maps (which were prepared three times daily using red and blue color markings to show the locations of German and Allied units on topographical maps of the area).

(2) Other sources included unit histories, official histories, and chronologies. Use was also made of the manuscripts prepared by German division, corps, and army commanders as part of the US Army's postwar Foreign Military Studies program.

(3) Daily unit location data was generated using the same general approach for all three nations. First the daily locations of divisions and brigades were determined. This information was then reviewed and their front line traces compiled to generate the data on corps. The process was repeated

using corps data for armies. Finally, the front line traces and locations of other significant unit elements of all forces were checked by comparing the daily front line locations of opposing Allied and German units. This step was necessary because the unit location data of opposing forces was derived separately. The proofing process involved plotting grid coordinates of units using the graphics feature of the ReflexTM data base management software. This check showed remarkably close correlation between the front line traces of the opposing forces, testimony to the accuracy of the German maps, the Allied situation reports, and the postwar manuscripts of the German officers.

e. Diskette Format.

(1) The data base is provided on a set of 33 5.25-inch 1.2 megabyte (high density) diskettes usable on IBM-compatible personal computers operating under MS-DOS 3.3 or higher, in a format readable by the dBASE IVTM data base management system. The ACSDB is divided into eight major data base files, each containing all the pertinent data on a particular aspect of the Ardennes Campaign. The eight major files are the: (i) Unit Data File, (ii) Unit Inventory File, (iii) Unit Location File, (iv) Air Data File, (v) Table of Organization and Equipment (TOE) File, (vi) Weapons Data File, (vii) Reference Data File, and (viii) Bibliography File. These files are composed of records ranging from a minimum of 153 (Bibliography File) to a maximum of 15,033 (Unit Inventory File) records per data base. Each record is composed of data fields, which contain identifying, statistical, or narrative information (such as a unit's name, number of infantry personnel, or a description of its operational mission). Space is too limited to describe the exact format of each file. However, the following general descriptions are provided.

(2) The Unit Data File contains personnel, medical, and logistical statistics of the American, German, and British ground combat units that participated significantly in the Ardennes Campaign. The fields in this file also record the order of battle for these units.

(3) The Unit Inventory Data File contains data on equipment (vehicles and weapons) strengths and losses for US, German, and British ground combat units. Equipment designation abbreviations used in this file are the same as those used in the Table of Organization and Equipment (TOE) and Weapons Data Files.

(4) The Unit Location File records information on the location of ground combat units, including independent brigades, divisions, corps, and armies. Unit locations are given in the form of map coordinates on the maps used by US and British forces at the time (which were original or US copies of the British "Nord de Guerre Zone," based on the Lambert Conformal Conical Projection, which the British call the Lambert Conical Orthomorphic Projection and described in the AMS Technical Manual No. 36, *Grids and Grid Projections*, Army Map Service, Corps of Engineers, Department of the Army, January 1950). This file also includes narrative descriptions of unit activities, operations, missions, and commander status for ground combat units.

(5) The Air Data File contains information on tactical air sorties flown by US Air Force, Royal Air Force, and Luftwaffe air units in the Ardennes battle area. This information includes the number of sorties, a description of operations and events during the missions, a geographical location of missions and air bases of origin of aircraft. Mission locations utilize the same map coordinates employed in the Unit Location File.

(6) The Table of Organization and Equipment (TOE) File gives information from the official TOE for all US, British, and German battalion-, regiment-, brigade-, and division-sized units covered in the ACSDB, plus some company-, troop-, and battery-sized units. Authorized personnel, equipment, and logistics strengths are provided for each unit in the TOE File. The same equipment designations used in the Unit Inventory and Weapons Data Files are used here.

(7) The Weapons Data File provides information on equipment (vehicles and weapons) characteristics. There are records in this file for air and ground systems tracked in the Inventory and TOE Files, including aircraft, armored fighting vehicles, indirect fire weapons, and soft-skinned transport vehicles. Weapons characteristics include rated movement rates of vehicles and aircraft, rated rate of fire, and ammunition loads of artillery and tank guns; weights of systems; *etc.* The same armament designations used in the Unit Inventory and TOE files are used here.

(8) The Reference Data File contains the sources used for data and information in other data base files. Each record in the Unit Location, Air, TOE, and Weapons Data Files contains a field identified as "Sources." Entries in those fields match identically an entry in the field "Source Number" of a record in the Reference Data File that identifies the sources consulted for data and information in the corresponding data file record.

(9) The Bibliographic Data File provides the full bibliographic citations for primary and secondary sources used for data file entries. The field "Bibliography Name" in a Reference Data File record is keyed to an entry under "Bibliography Name" in the Bibliographic Data File, thus completing the connection between the "Sources" entry in a data file and the full bibliography citation supporting the corresponding data items. The written narrative in the hard copy documentation provide additional information supplementing the Reference Data and Bibliographic Data files and facilitating comprehensive referencing of all data in the ACSDB.

3. Other and Miscellaneous. Examples of the use of this data base may be found in the following studies:

a. "A Compilation of Data on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-4, February 1990 (UNCLASSIFIED, available from DTIC as AD-A220 426).

b. "A Survey of Past Work on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-3, February 1990 (UNCLASSIFIED, available from DTIC as AD-A217 891).

c. "Rates of Advance in Historical Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-1, June 1990 (UNCLASSIFIED, available from DTIC as AD-A225 635).

4. Remarks.

a. This is a major effort to produce a comprehensive quantitative record of the events that occurred during the Ardennes Campaign. The data are in a form that lends itself to ready retrieval and collation.

b. Since the data were entered by several individuals at many different times, it was not possible to stick exactly to the stated data base format. For example, missing values may be variously coded as "99999999," "999.9," or as a blank field with no entries. This must be taken into account when using computer programs to process the data.

c. Some data are not coded in a manner that facilitates computer processing. For example, the same weather conditions may be variously described as "fog," "foggy," "limited visibility," "visibility limited to 100-200 meters," "moderately foggy," "fog obscured enemy positions," "visibility of enemy positions limited," "visibility hampered operations," etc. Computer processing of the information in these fields is impractical in their present format. With enough additional work, it might be possible to convert the information to a standard format.

e. There is no direct record in the data base as to which of the distances described in the hard copy documentation were actually entered into fields 7 (Displacement) and 8 (Distance Opposed Advance) of the Unit Location File. To some extent that information can be inferred from the context, but not in a way that is conveniently programmed for computer manipulation of the information.

ADE-1992

1. **General Description.** Helmbold, Robert L., "Ardennes Division Engagement Data," US Army Concepts Analysis Agency unpublished material, 1992, UNCLASSIFIED. Diskette available from Helmbold as ADE.WQ!.

2. **Data Base Description.**

a. **General.** The diskette contains one file describing personnel losses to selected divisions and independent brigades, and to their corresponding opponents, derived from the Ardennes Campaign Simulation Data Base (ACSDB, available from DTIC as AD-M000 074). The following is a simplified description of the procedure by which the ADE.WQ! file was constructed.

Although the ACSDB gives information on losses to individual units, it provides no easy way to determine the mutual losses of units engaged with each other. In order to develop this sort of information for a sample of engagements by divisions and independent brigades, the following procedure was used. In essence, each division and independent brigade listed in the ACSDB data was considered as a "base unit." By considering the dispositions of the units, as recorded in the ACSDB's Unit Location data files, an attempt was made to match this base unit to the specific divisions or independent brigades with which it was engaged on a particular day. If the divisions or independent brigades opposing the potential base unit on that day could clearly be identified, then it was promoted to a "successful matchup." In this manner, a total of 393 successful matchups was obtained.

The total personnel strength of the force opposing the base unit was estimated using the proportion of frontage each opposing unit shared with the base unit. For example, suppose that some base unit's frontage was 40 km, that it was opposed by two enemy units whose personnel strengths and frontages were 30,000 on a front of 30 km and 40,000 on a front of 20 km, and that the fronts of these enemy units overlapped that of the base unit by 25 km and 15 km, respectively. Then the personnel strength opposing the base unit was estimated as illustrated by the following computation:

$$25 \text{ km} \times 30,000 \text{ men}/30 \text{ km} + 15 \text{ km} \times 40,000 \text{ men}/20 \text{ km} = 55,000 \text{ men.}$$

The opposing unit's casualties and other changes in personnel strength were also apportioned to the base unit's frontage in the same fashion.

b. **Information Provided.** For each of the 393 base units, the data base gives the following information:

Seq. No. (an arbitrary sequence number to number the rows)
 Day (from day 1 = 16 December 1944 to day 32 = 16 January 1945)
 Date (the date corresponding to the Day)
 Side (A = American, G = German, B = British)
 Base Unit ID (name of the base unit, e.g., 106th ID)
 German Initial Strength, x0

Allied Initial Strength, y0
 German Total Battle Casualties, Cx
 Allied Total Battle Casualties, Cy
 German Bloody Casualties, Bx
 Allied Bloody Casualties, By
 German Other Changes, Rx
 Allied Other Changes, Ry
 Distance Opposed Advance by Allies (km)

Here, strengths, losses, and other changes are for personnel, and (as explained earlier) are assumed to be in proportion to the overlap of frontages. Initial strengths are those at the start of the day. Casualties and other changes are those that occurred during the day. Total battle casualties include KIA, WIA, MIA, and POW. Bloody casualties include only KIA and WIA. Other changes are those not due to battle casualties (*e.g.*, DNBI, reinforcements, returns to duty, administrative attachments and detachments, *etc.*).

The distance advanced is the distance the front moved during the day. Distances are expressed from the viewpoint of the Allies. That is, a negative value indicates that the Allies were moving backward, and a positive value indicates they were moving forward. These values are based on the "Distance Opposed Advance" recorded in the ACSDB files for the base unit.

c. **Situational Descriptors Included.** Not used.

d. **Data Sources Used.** "The Ardennes Campaign Simulation Data Base," Data Memory Systems, Inc., Fairfax, VA, 18 December 1989, UNCLASSIFIED, available from DTIC as AD-A240 088 (documentation) or as AD-M000 074 (data diskettes, including documentation).

e. **Diskette Format.** The ADE diskette contains one file: ADE.WQ!, which is in spreadsheet form with rows for the individual division engagements and columns for the items of information provided on each engagement.

3. **Other and Miscellaneous.** Not used.

4. **Remarks.** The opposing strengths and losses are assumed to be in proportion to the opposing unit's frontages, although this is clearly not always the case. Also, in the ACSDB, due to lack of specific historical data, the strength and casualty figures for many individual units had to be estimated. In view of these considerations, the strength and loss data values carried over to the ADE data set may be in error for specific units on specific days. For example, although a given base unit's casualty experience can frequently be followed for several consecutive days, its starting strength on any given day often differs from what would be expected from its previous day's initial strength, losses, and gains.

AHART-1993

1. **General Description.** Dougherty, Francis L., "Analysis of Historical Artillery Expenditures (AHART) Study-CY87." US Army Concepts Analysis Agency, Technical Paper CAA-TP-87-6, June 1987, UNCLASSIFIED. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Contains 3,594 records of artillery expenditures by selected US units from World War II through the Vietnam War. Of these, 950 records have both (i) a number of rounds per day that is greater than zero, and (ii) a number of days greater than one.

b. **Information Provided.** Information in each record indicates (to the extent known) the source, battle, supported unit, size of supported unit, date, artillery unit, tube type, tube category, tube quantity, type round, rounds per day, round quantity, number of days, operation, round weight, round weight per day, total weight, round per tube per day, round cost, round cost per day, total cost, notes, and some additional items whose values are seldom known. No casualty data are provided.

c. **Situational Descriptors Included.** None, other than the name and date of the engagement

d. **Data Sources Used.** Various, but in all cases were secondary sources.

e. **Diskette Format.** The data is provided in ASCII, as a "flat file" of 3,594 records. Each record is 342 characters long (not counting the carriage return/line feed ASCII characters that terminate each line). The fields within each record (many of which contain blanks or other codes indicating that the value is not known) are laid out in the following fashion.

<u>Field Name</u>	<u>Field Width in Characters</u>
Source	20
Battle	12
Unit	12
Size	8
Date	8
ArtyUnit	12
TubeType	11
TubeCat	7
TubeQuant	12
TypeRd	6
RdPerDay	10
RoundQuant	8
DaysQuant	3
Operation	9
RoundWt	5
RdWtDay	10
TotalWt	12
RdTubeDay	6
RoundCost	4
RdCostDay	10
TotalCost	12

Notes	10
RedUnit	12
RSize	8
RedOpn	10
RTypeTube	12
RTubeQuant	10
RTypeRd	6
RRdPerDay	10
RRdQuant	12
Misc	<u>57</u>
Total	344

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following studies:

a. US Army Concepts Analysis Agency Technical Paper CAA-TP-87-6, "Analysis of Historical Artillery Expenditures (AHART) Study-CY87," June 1987.

b. US Army Concepts Analysis Agency, Memorandum Report CAA-MR-92-32, "Benchmark for Artillery Munitions Consumption (BAMC)," May 1992.

4. **Remarks.**

a. This is probably the largest collection of artillery ammunition expenditure data readily available on diskette.

b. However, this data base is not the same as documented in the report cited in paragraph 1 above. Instead, it was recreated from electronic files that (although based largely on that report) had suffered several transmutations of form and substance. Accordingly, while the information in the AHART-1993 files is believed to be reasonably correct, it should be rechecked before using it as the basis for important decisions. Since this data base clearly has been modified in various ways (and for uncertain reasons) since MAJ Dougherty's original compilation, it must be used circumspectly.

BENCHMARK-1988

1. **General Description.** McQuie, Robert, "Data on Historical Battles 1937-1982 (Benchmarks)," US Army Concepts Analysis Agency Research Paper CAA-RP-87-2, July 1988, 60 pp, UNCLASSIFIED. Report available from DTIC as AD-A200 036. Diskette available from DTIC as AD-M000 000 or AD-M000 092 (one diskette, UNCLASSIFIED).

2. **Data Base Description.**

a. **General.** Provides data on both sides in 257 battles that took place from 1937 through 1982. Selected battles are included from World War II, the Korean War, and the Arab-Israeli War of 1948, 1956, 1967, 1973, and 1982. Two-thirds of the battles took place in World War II, one-third since 1945. The records on almost all of them are incomplete, and one-fifth of the cells in the spreadsheet are blank.

b. **Information Provided.** Information on each battle includes the battle name, location, theater, and date; the opposing force nationalities and echelons (*e.g.*, division, regiment, or corps); terrain (rolling, rugged, flat); cover (woods, bare, urban, swamp, dunes, mixed); humidity [precipitation] (dry, light, heavy); temperature (hot, temperate, cold); attacker's maneuver, width of front, depth, and surprise; defender's posture [fortifications], width of front and depth of position; duration in days and hours; personnel and tank casualties to the attacker and defender; movement [distance advanced]; resolution [outcome] for the attacker and defender; and the attacker's and defender's forces in men, rifle squads, mortars, antitank weapons, tanks, air defense weapons, artillery (guns and missiles), and close air support sorties. The casualty category used is total battle casualties (the sum of the KIA, WIA, MIA, and POW).

c. **Situational Descriptors Included.** As stated under primary information included.

d. **Data Sources Used.** Various, but mostly from the same sources as were used in the CDB90 data base.

e. **Diskette Format.**

(1) The data base is provided on one 5.25-inch 360 kilobyte (low density) diskette usable on IBM-compatible personal computers operating under MS-DOS 3.21 or higher.

(2) It is also provided as a 3.5-inch 1.75 megabyte (high density) diskette usable on IBM-compatible personal computers operating under MS-DOS 3.21 or higher as DTIC AD-M000 092.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the US Army Concepts Analysis Agency Research Paper CAA-RP-87-2, "Historical Characteristics of Combat for Wargames (Benchmarks)," July 1988, available from DTIC as AD-A200 036.

4. **Remarks.** Most of the data in this data base has been incorporated into the more extensive CDB90-1990 data base, which as a result largely supersedes the Benchmark-1988 data base.

BERNDT-1993

1. **General Description.** Taken from Berndt, Otto (Captain in the Austrian General Staff), *Die Zahl im Kriege: Statistische Daten aus der Neueren Kriegsgeschichte in Graphischer Darstellung* [Number in War: Statistical Data from Modern Military History in Graphical Form], G. Freytag & Berndt, Vienna, 1897, 169 pp. UNCLASSIFIED. Available from US Army Command and General Staff College Library (355.09 B524z). Data base on diskette furnished to DTIC concurrently with this report.

2. **Data Base Description.**

a. **General.** Contains data on the strengths and losses to each side in 78 European battles from Mollwitz (10 April 1744) to Plevna (10 December 1877).

b. **Information Provided.** Information on each battle includes the battle name; the year, month and day on which the battle began; identification of the principal nationality for the victorious and losing side; initial strengths of the two sides; bloody and CMIA losses to each side; and duration of the battle in hours. The casualty categories used are bloody losses (KIA plus WIA), and CMIA.

c. **Situational Descriptors Included.** None, other than the name and date of the engagement.

d. **Data Sources Used.** Various, but include unclassified Austrian General Staff publications.

e. **Diskette Format.** One file in spreadsheet form with rows for the individual battles and columns for the items of information provided on each.

3. **Other and Miscellaneous.** An example of the use of this data base may be found in Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group (CORG), HQ, US Continental Army Command, Staff Paper CORG-SP-128, 1 July 1961, 178 pp, UNCLASSIFIED. Available from DTIC (AD-480 975).

4. **Remarks.**

a. Curiously, Berndt did not supply an identification of which side was the attacker and which the defender. Bodart, in his *Kriegslexicon*, followed this same practice.

b. The data base omits the battle of Schlact-bei-Noiseville listed in Berndt's tabulation. It was omitted because it was a 2-day battle with substantial changes in the forces engaged between the first and second days, but Berndt does not give enough information to determine how many of the total losses were suffered on each day. Thus it is not possible to determine the proper initial forces and their losses for either day.

BOB18-1993

1. **General Description.** Helmbold, Robert L., "Air Battles and Ground Battles—A Common Pattern?", RAND Corporation Paper P-4548, 1971, UNCLASSIFIED, available from DTIC as AD-718 975. Data base on diskette furnished to DTIC concurrently with this report.
2. **Data Base Description.**
 - a. **General.** Contains Luftwaffe and Fighter Command sorties and losses for each of the 18 days of the Battle of Britain (13 August 1940 to 6 September 1940).
 - b. **Information Provided.** Information on each day includes the date; Luftwaffe sorties; Fighter Command sorties; and aircraft lost to each side. The casualty category used is aircraft losses.
 - c. **Situational Descriptors Included.** None, other than the name and year of the engagement.
 - d. **Data Sources Used.** The data were taken from Collier, B., *Defense of the United Kingdom,* Her Majesty's Stationery Office, London, 1957, UNCLASSIFIED, copies obtainable from libraries.
 - e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 18 days and columns for each of the items of data provided.
3. **Other and Miscellaneous.** An example of the use of this data base may be found in the RAND paper cited in paragraph 1 above.
4. **Remarks.** Does not deal with land combat forces.

BODASHIP-1993

1. **General Description.** A collection of naval battle data taken from Bodart, Gaston, *Militar-Historisches Kriegs-Lexicon (1618-1905)*, C. W. Stern, Vienna and Leipzig, 1908, 956 pp. UNCLASSIFIED, available from libraries. Data base on diskette furnished to DTIC concurrently with this report.
2. **Data Base Description.**
 - a. **General.** The diskette contains one file describing losses to both sides in the naval battles documented in the source.
 - b. **Information Provided.** For each of 120 naval battles, the data base gives Bodart's page and item on the page from which the data were taken, the type of battle (Seeschlacht, Seetreffen, or Seegefecht), the battle name, starting and ending dates, duration of the battle (in whole days), the principal nations on the winner's and on the loser's side, and the following information for the winner and loser: (i) the initial status (personnel strength, number of main ships, number of other ships, and number of cannon), (ii) losses (dead, wounded, bloody, MIA/POW, total personnel losses, of main ships, of other ships, and of cannon).
 - c. **Situational Descriptors Included.** None, other than the name and date of the battle.
 - d. **Data Sources Used.** Various.
 - e. **Diskette Format.** The diskette contains one file, which is in spreadsheet form with rows for the individual battles and columns for the items of information provided on each battle.
3. **Other and Miscellaneous.** Not used.
4. **Remarks.** Although not directly related to land combat operations, this data base of naval battles is included as it appears to be unique in its compilation of naval engagement data.

BWSH-1993

1. **General Description.** Bodart, Gaston, *Militär-Historisches Kriegs-Lexicon (1618-1905)*, C. W. Stern, Vienna and Leipzig, 1908, 956 pp, UNCLASSIFIED, available from libraries. Data base on diskette furnished to DTIC concurrently with this report.

2. **Data Base Description.**

a. **General.** Includes strength and loss data on 1,087 land combat battles that occurred between 1619 and 1905. The intent was to include all land combat battles for which Bodart's *Kriegslexicon* gives strengths and losses to both sides. The data base omits some of Bodart's battles (mainly sieges) in which disease and/or reinforcements caused marked changes in the opposing strengths at unknown times during the engagement.

b. **Information Provided.** The information on each battle includes Bodart's page and item number on the page; the type of battle according to Bodart's terminology (i.e., Treffen (engagement), Gefecht (combat), Schlacht (battle), Ersturmung (assault or storming of fortifications), or Belagerung (siege)); the battle name; its starting and ending year, month, and day; its duration in whole days; identification of the winner's and the loser's primary nationalities (at most two nations are listed for each side—e.g., France and Austria might be listed as victors over Prussia and Poland, even though smaller forces may have been contributed by other nationalities on each side); the total strength committed to the battle by each side; and the losses on each side. The casualty categories used are dead, wounded, bloody losses (i.e., the sum of the dead and wounded), bloodless losses (i.e., the missing and captured, lumped into a single figure), and total battle casualties (i.e., the sum of the bloody and bloodless losses).

c. **Situational Descriptors Included.** None, other than the name and year of the engagement.

d. **Data Sources Used.** The history of this data base is as follows. The original data were published by the famous military historian Gaston Bodart in his monograph cited in paragraph 1 above. The land combat battle data it contains were encoded for electronic data processing at the Research Analysis Corporation in the early 1960s to support Dr. Daniel Willard's analyses (see paragraph 3b, below). In the mid-1960s, Dr. Willard provided the data deck to Dr. William A. Schmieman, then a graduate student at the Georgia Institute of Technology, who used them in his thesis research (see paragraph 3d, below). In the late 1960s, Dr. Schmieman kindly provided a copy of his version of the data base to Dr. Robert L. Helmbold, who retained it for several years and made use of it from time to time in his research (see paragraphs 3e, f, and g, below). In the 1980s Dr. Robert L. Helmbold converted the card deck to magnetic tape, and subsequently to diskettes usable on personal computers. Then in 1993, Dr. Robert L. Helmbold carefully edited and revised the data base by comparing it item by item to Bodart's original, correcting several errors, revising the data format to

enhance its clarity, including additional information, and eliminating redundancies. The additional information included at that time was mainly the division of bloody losses into the dead and wounded whenever Bodart's tabulation provided these values. To give one example of the elimination of redundancies, we note that Bodart himself categorized battles by magnitude according to the total number of personnel involved and that these were carried into the card decks used by Willard and Schmieman. However, retaining these magnitude codes is redundant because in this age of economical computation it is simpler and easier to compute Bodart's categories anew as and if they are needed instead of embedding them in the data base. For example, this facilitates flexibly redefining such categories to conform to the needs of particular analyses, rather than sticking with rigidly predetermined categories.

c. Diskette Format. The data are given in three spreadsheet files; BWSH1600 (Bodart land combat battles from 1619 through 1699), BWSH1700 (Bodart land combat battles from 1700 through 1799), and BWSH1800 (Bodart land combat battles from 1800 through 1905). Each spreadsheet has rows for the battles and columns for the items of information provided on each battle.

3. Other and Miscellaneous. Examples of the use of Bodart's data may be found in at least the following studies:

a. Wright, Quincy, *A Study of War*, University of Chicago, 1942 (revised edition 1965), (Vol I is 678 pp; Vol II is 873 pp), UNCLASSIFIED, available from CAA Library as 355.02 WRI.

b. Smith, E. G.; and Donovan, G. N., "Trends in Warfare," British Army Operational Research Group, Memorandum No. F.6, April 1955.

c. Willard, Daniel, "Lanchester as Force in History: An Analysis of Land Battles of the Years 1618-1905," Research Analysis Corporation Technical Paper RAC-TP-74, November 1962 (AD-297 375L).

d. Klingberg, Frank L., "Predicting the Termination of War: Battle Casualties and Population Losses," *Journal of Conflict Resolution*, Vol X, No 2, 1966, pp 129-171.

e. Schmieman, William A., "The Use of Lanchester-Type Equations in the Analysis of Past Military Engagements," Doctoral Thesis, School of Industrial Engineering, Georgia Institute of Technology, August 1967.

f. Helmbold, Robert L., "Probability of Victory in Land Combat as Related to Force Ratio," The RAND Corporation, RAND P-4199, October 1969. AD-696 489.

g. Helmbold, Robert L. "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," The RAND Corporation, RAND R-772-PR, June 1971. AD-729 769.

h. Helmbold, Robert L. (et al), "Integrated Warfare—Representing the Decision Processes Within Computer Simulations of Combat," US Army Concepts Analysis Agency Technical Paper CAA-TP-85-1, January 1985.

4. **Remarks.** Bodart's *Kriegslexicon* does not indicate which side was attacking or defending, so no information on that is provided in the BWSH data bases. Willard's paper claims 1,088 battles with strengths and losses given for both sides, Schmieman's thesis states that there were 1,081 battles in the data base he used, the version that came into Dr. Robert L. Helmbold's hands contained exactly 1,080 battles, and the "revised standard" version described above contains 1,087 battles. I have no explanation for these discrepancies except for the last, which came about when a few items were added or dropped as errors were corrected by carefully comparing the data base to Bodart's original.

1. **General Description.** Helmbold, Robert L., "CAA's Data Base of Battles—Version 1990," US Army Concepts Analysis Agency, unnumbered, UNCLASSIFIED. For description of the documentation, see paragraph 2e (Data Sources Used), below. Diskette available from DTIC as AD-M000 121 (two diskettes, UNCLASSIFIED, 30 April 1991).

2. **Data Base Description.**

a. **General.** Includes 660 battles that were fought between 1600 AD and 1973 AD. Descriptive data include battle name, date, and location; the strengths and losses on each side; identification of the victor; temporal duration of the battle; and selected environmental and tactical environment descriptors (such as type of fortifications, type of tactical scheme, weather conditions, width of front, etc).

b. **Information Provided.** Information on each battle includes the following: battle sequence (an arbitrary index or serial number); war in which battle/engagement was fought; name of the battle/engagement; location where battle was fought (country or region); campaign in which battle was fought; designation of the attacking forces; name of the attacking force commander; designation of the defending forces; name of the defending force commander; attacker's width of front at start of battle; defender's width of front at start of battle; time first widths of front became effective; attacker's second width of front; defender's second width of front; time second widths of front became effective; attacker's third width of front; defender's third width of front; time third widths of front became effective; defender's posture type descriptor [0 = At most one defensive posture is given, 1 = A combination posture involving two distinct/separate postures, 2 = An average or intermixture of two or more postures, 9 = More than one posture, but whether separate or intermixed is not available]; defender's primary defensive posture [HD = Hasty defense, PD = Prepared defense, FD = Fortified defense, DL = Delaying action adopted, WD = Withdrawal adopted, OO = Not available or not applicable]; defender's secondary posture; did defender's posture change along the front? [0 = No, 1 = Yes, 9 = Unknown]; did defender's posture change with depth? [0 = No, 1 = Yes, 9 = Unknown]; did defender's posture change over time? [0 = No, 1 = Yes, 9 = Unknown]; primary local terrain description [First character: G = Rugged, R = Rolling, F = Flat, 0 = Other or not available; second character: W = Heavily Wooded, M = Mixed, B = Bare, D = Desert, 0 = Other or not available; third character: U = Urban, M = Marsh or swamp, D = Dunes, 0 = Other or not available]; secondary local terrain description; primary local weather description [First character: W = Wet, D = Dry, 0 = Other or not available; second character: H = Heavy Precipitation, L = Light Precipitation, O = Overcast (no precipitation), S = Sunny (no precipitation), 0 = Other or not available; third character (for local weather): H = Hot, T = Temperate, C = Cold, 0 = Other or not available; fourth character (for general

climate); E = Tropical (i.e., "Equatorial"), D = Desert, T = Temperate, 0 = Other or not available]; secondary local weather description; tertiary local weather description; relative surprise achieved by attacker [+3 = Complete surprise achieved by attacker, +2 = Substantial surprise achieved by attacker, +1 = Minor surprise achieved by attacker, 0 = Neither side achieved surprise, or it did not affect the outcome, -1 = Minor surprise achieved by defender, -2 = Substantial surprise achieved by defender, -3 = Complete surprise achieved by defender, 9 = Unknown or not available]; attacker's relative air superiority in theater [+1 = Attacker had air superiority in the theater, 0 = Neither side had theater air superiority, -1 = Defender had air superiority in the theater, 9 = Unknown or not available]; attacker's total personnel strength; interpretation of attacker's total personnel strength [1 = Total strength is also initial strength, 2 = Total strength is an average daily strength, 3 = Total strength is the total strength engaged]; attacker's initial personnel strength; attacker's personnel reinforcements/replacements during battle; attacker's personnel battle casualties; attacker's personnel strength at battle's end; defender's total personnel strength; interpretation of defender's total personnel strength; defender's initial personnel strength; defender's personnel reinforcements/replacements during battle; defender's personnel battle casualties; defender's personnel strength at battle's end; attacker's horse cavalry strength; attacker's total number of tanks; attacker's number of light tanks; attacker's number of main battle tanks; attacker's number of artillery tubes; number of close air support sorties by attacker; attacker's armor losses; attacker's artillery losses; attacker's close air support aircraft losses; defender's horse cavalry strength; defender's total number of tanks; defender's number of light tanks; defender's number of main battle tanks; defender's number of artillery tubes; number of close air support sorties by defender; defender's armor losses; defender's artillery losses; defender's close air support aircraft losses; attacker's relative combat effectiveness;¹ attacker's relative leadership advantage;¹ attacker's relative training advantage;¹ attacker's relative morale advantage;¹ attacker's relative logistics advantage;¹ attacker's relative momentum advantage;¹ attacker's relative intelligence advantage;¹ attacker's relative technology advantage; attacker's relative initiative advantage;¹ attacker's win/draw/lose [+1 = ATK Win, -1 = ATK Loss, 0 = Draw, -9 = Unknown]; distance attacker advanced/retreated; attacker's mission accomplishment score [on scale of 1 to 10, 1 = Low, 10 = High, 0 = Unknown]; defender's mission accomplishment score; criterion for assigning attacker's win/draw/lose [1 if clear-cut, 2 if not, 0 = not available]; degree of influence of force quality;¹ degree of influence of reserves;¹ degree of influence of mobility;¹ degree of influence of air superiority;¹ degree of influence of force preponderance;¹ degree of influence of weather;¹ degree of influence of terrain;¹ degree of influence of leadership;¹ degree of influence of planning;¹ degree of influence of surprise;¹ degree of influence of maneuverability;¹ degree of influence of logistics;¹ degree of influence of fortifications;¹

¹ +4 = Very strongly favors the attacker, +3 = Strongly favors the attacker, +2 = Favors the attacker, +1 = Somewhat favors the attacker, 0 = Favors neither side, -1 = Somewhat favors the defender, -2 = Favors the defender, -3 = Strongly favors the defender, -4 = Very strongly favors the defender, -9 = Unknown or not available.

degree of influence of depth of position;¹ attacker's primary tactical scheme: part 1 [FF = Frontal attack; EE = Single envelopment; DE = Double envelopment; FE = Feint, demonstration, or holding attack; DD = Defensive plan; DO = Defensive/offensive plan; LF = Left flank; RF = Right flank; LR = Left rear; RR = Right rear; PP = Penetration; RC = River crossing; 00 = Not applicable, or not available]; attacker's primary tactical scheme: part 2; attacker's primary tactical scheme: part 3; attacker's secondary tactical scheme: part 1; attacker's secondary tactical scheme: part 2; attacker's secondary tactical scheme: part 3; attacker's primary resolution/outcome: part 1 [AA = Annihilated, PS = Pursued, WL = Withdrew with heavy losses, WD = Withdrew, BB = Breakthrough, PP = Penetration, RR = Repulse, SS = Stalemate, 00 = None of the above, or not available]; attacker's primary resolution/outcome: part 2; attacker's primary resolution/outcome: part 3; defender's primary tactical scheme: part 1; defender's primary tactical scheme: part 2; defender's primary tactical scheme: part 3; defender's secondary tactical scheme: part 1; defender's secondary tactical scheme: part 2; defender's secondary tactical scheme: part 3; defender's primary resolution/outcome: part 1; defender's primary resolution/outcome: part 2; defender's primary resolution/outcome: part 3; attacker strength error [\pm percent of tabulated value]; attacker strength error; attacker casualty error; attacker casualty error; defender strength error; defender strength error; defender casualty error; defender casualty error; number of active time periods for this battle [ranges from 1 to 10]; time first active time period began/ended; time second active time period began/ended; *etc.* The casualty category used is total battle casualties (the sum of the KIA, WIA, MIA, and POW).

c. **Situational Descriptors Included.** As described above in paragraph 2b.

d. **Data Sources Used.** The history of this data base is summarized here. CAA's data base of information on historical land combat battles and engagements exists in two distinct forms. One is the traditional printed report form, and the other is in the form of computerized data files. Both versions contain historical information on over 600 land combat battles that took place between 1600 AD and 1990 AD. The report version has gone through several iterations, as described below.

(1) The first version of the data base is documented in CAA Study Report CAA-SR-84-6, "Analysis of Factors That Have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements," September 1984, in six volumes. It was prepared for CAA under contract number MDA903-82-C-0363, and is available from DTIC under the following accession numbers: AD-B086-797L, AD-B087-718L, AD-B087-719L, AD-B087-720L, AD-B087-721L, and AD-B087-722L.

(2) The second version is documented as "changes to" the first version in HERO Report Number 129, "Combat History Analysis Study Effort (CHASE) Data Enhancement Study (CDES)," 31 January 1986, in five volumes. It was prepared for CAA under Contract Number MDA903-85-C-0252 and is available from DTIC under the following accession numbers: AD-A175-712, AD-A175-713, AD-A175-714, AD-A175-715, and AD-A175-716. The objective of the CDES work was to correct omissions,

inconsistencies, and ambiguities in the battle and engagement data base being used in the CHASE Study. Its scope included: (i) analyze about 400 data base problem reports generated while encoding the original HERO data into machine-readable format; (ii) state for each battle whether its total engaged personnel strength is the initial, average, or total strength; (iii) state for each battle whether HERO determined its victorious side on the basis of a clear-cut decisive resolution, or on the basis of mission accomplishment; (iv) give for each battle duration data in hours and minutes rather than in "days;" (v) state for each battle the defender's width of front; (vi) state for each battle with a dual posture descriptor whether it represents an "average" or a "combination" posture; (vii) provide for each battle estimates of the relative reliability of its personnel strength and casualty data; (viii) develop strength and attrition histories for selected battles; (ix) comment on CAA's attempt to eliminate unwanted redundancies in the data. The principal findings of CDES are that, despite the care taken in the data base contract to prepare the original data base, and despite the further effort in the CDES contract to rid it of errors, a number still remain. Lessons learned included the practical impossibility of assuring the high reliability of historical data bases.

(3) The third version is documented as "changes to" the first and second versions in unnumbered HERO Report, "Data Base Error Correction (DBEC)," 23 January 1987. It was prepared for CAA under Purchase Order Number MDA903-86-M-8560 and is available from DTIC under accession number AD-A176-750. The DBEC effort was motivated by a desire to purge the data base of as many errors as humanly possible. Its scope included: (i) correct advance rate data; (ii) resolve discrepancies in the duration data; (iii) correct a variety of errors; (iv) provide definitions for two data types that were previously undefined; (v) provide a list of errata for the CDES report. The principal finding of the DBEC effort was that discrepancies remain between some of the narrative descriptions of battle and the data base—no attempt was made to correct them within the scope of the DBEC effort. The major topic for future research is that not all of the errors in the data base have as yet been identified and corrected. The lessons learned underscored the difficulty of assuring high reliability in historical combat data bases.

(4) In addition to the above, an independent review and reassessment of the data for about 60 selected battles whose data are considered to be statistically anomalous is documented in the unnumbered LFW Management Associates, Inc. report "Independent Review/Reassessment of Anomalous Data (IR/RAD)," unnumbered LFW Management Associates, Inc. Report, 22 June 1987, in four volumes. Prepared for CAA under Contract Number MDA903-86-C-0396, and available from DTIC under accession numbers AD-195 381 (Volume I), AD-195 725 (Volume II), AD-195 382 (Volume III), and AD-195 383 (Volume IV). The objective of the IR/RAD work was to perform an independent review/reassessment of certain historical battle data. It was motivated by the CHASE Interim Progress Report's finding that the data for these particular battles are statistically anomalous. The scope of the IR/RAD effort included developing new and original historical data for 61 anomalous

battles in 8 campaigns of World War II and the Arab-Israeli wars. The aim was to provide a factual basis for determining whether the anomalies are attributable to actual changes in combat dynamics, or whether they are more likely due to flaws in the data base.

The principal finding was that, in virtually every case, the IR/RAD contractor's data differ substantially from those determined by the authors of the original data base. Major topics for future research include determining the validity, exact nature and extent, theoretical implications, and practical significance of these differences between the two groups of historical analysts. The most important lesson learned appears to be that it is very difficult to assure high quality in historical combat data bases.

(5) Use was also made of Data Memory Systems Incorporated report "New Engagement Data for the Breakpoints Data Base," prepared for the US Army Concepts Analysis Agency under Contract No. MDA903-87-C-0807, 30 September 1988. That report, in conjunction with independent research by Mr. Robert McQuie of CAA, provided data on about 59 additional battles, and these are included in CAA's land combat data base, Version 1990 (CDB90).

e. **Diskette Format.** The data base is provided on a set of two 5.25-inch 1.2 megabyte (high density) diskettes usable on IBM-compatible personal computers operating under MS-DOS 3.21 or higher. The files on these diskettes are as follows:

- (1) README.TXT, an ASCII file describing the data base.
- (2) CDB90DEF.WKS, a LOTUS-readable spreadsheet file containing definitions of the variables and coding scheme used in the data base.
- (3) CDB90001.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 1 through 74.
- (4) CDB90075.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 75 through 149.
- (5) CDB90150.WKS, LOTUS-readable spreadsheet file containing the data base information for battles 150 through 224.
- (6) CDB90225.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 225 through 299.
- (7) CDB90300.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 300 through 374.
- (8) CDB90375.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 375 through 449.

(9) CDB90450.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 450 through 524.

(10) CDB90525.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 525 through 599.

(11) CDB90600.WKS, a LOTUS-readable spreadsheet file containing the data base information for battles 600 through 660.

3. **Other and Miscellaneous.** Portions of the data base are known to have been used in at least the following studies, among others:

a. "Integrated Warfare—Representing the Decision Processes Within Computer Simulations of Combat," US Army Concepts Analysis Agency Technical Paper CAA-TP-85-1, January 1985, 175 pp.

b. "Combat History Analysis Study Effort (CHASE): Progress Report for the Period August 1984-June 1985," CAA-TP-85-2, August 86. Available from DTIC as accession number AD-A179 734.

c. "Do Battles and Wars Have a Common Relationship Between Casualties and Victory?," US Army Concepts Analysis Agency Technical Paper CAA-TP-87-15, November 1987. Available from DTIC as accession number AD-A196 126.

d. "Course Design: Osipov's Law," US Military Academy Cadet Project, 3 May 1988, 67 pp.

e. "Historical Characteristics of Combat for Wargames (Benchmarks)," US Army Concepts Analysis Agency Research Paper CAA-RP-87-2, July 1988. Available from DTIC as accession number AD-A200 036.

f. "Can the Square Law be Validated?," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-57, March 1989, 72 pp.

g. "Historical Support for a Mixed Law Lanchestrian Attrition Model: Helmbold's Ratio," Dean S. Hartley and K. L. Kruse, Data Systems Research & Development Program Report K/DSRD-113, November 1989, 51 pp.

h. "The Constraint Model of Attrition," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-114/R1, November 1989, 24 pp.

i. "A Compilation of Data on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-4, February 1990 (UNCLASSIFIED, available from DTIC as AD-A220 426).

j. "A Survey of Past Work on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-3, February 1990 (UNCLASSIFIED, available from DTIC as AD-A217 891).

k. "Rates of Advance in Historical Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-1, June 1990 (UNCLASSIFIED, available from DTIC as AD-A225 635).

l. "Historical Validation of an Attrition Model," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-115, May 1990, 51 pp.

m. "Variation of Historical Casualty Rates With Battle Dates," Cathy J. Arebalo, U.S. Army Soldier Support Center unnumbered paper, December 1990, 30 pp, available from DTIC as AD-A235 876.

n. "Confirming the Lanchestrian Linear-Logarithmic Model of Attrition," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-263/R1, February 1991, 136 pp.

o. "User's Guide to the Oak Ridge Spreadsheet Battle Model," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-413, May 1991, 55 pp.

p. "Technical Notes on Comparison of Battle Outcomes Across Nationality", Jean A. Bredeson, US Army TRADOC Analysis Command-Ft. Benjamin Harrison unnumbered paper, 29 May 1991, 51 pp.

q. "Predicting Combat Effects," Dean S. Hartley III, Data Systems Research & Development Program Report K/DSRD-412, August 1991, 360 pp.

r. "A Search for Chaotic Behavior in Battle Outcomes," CPT Singleton, Department of Mathematical Sciences, US Army Military Academy, West Point, 1992.

4. Remarks.

a. The CDB90 data base provides what is perhaps the most complete and detailed data base on battles that is currently available. The work described above has resulted in a comprehensive data base of land combat battles from 1600 AD to the present, describing quantitatively many of their features. Despite its shortcomings, it is the largest and most comprehensive free world data base of quantitative information on battles. It is particularly well suited to statistical analyses of the type described in paragraph 3 above. Its preparation was motivated by the understanding that historical battle data are essential for important advances in the quality of Army modeling, together with a recognition that the previously available combat data were either insufficiently detailed or not systematically organized into the quantitative data base form required for validating contemporary battalion to army-level models. The scope of the data base includes a tabulation of about 80 items of information and a synopsis of the action for each battle.

b. An important finding is that it is feasible to construct an extensive and well-organized data base of quantitative information on battles. Important issues and topics for future research include devising improved criteria for deciding what data to put in the data base, developing improved approaches to analyzing this sort of data, and inventing ingenious ways of applying the results to important Army issues. Among the important lessons learned are that:

(1) The successful development of a high-quality combat data base requires a multidisciplinary approach in which military archivists and historians work together with military operations analysts, statisticians, epistemologists, and scientists with a strong "hard science" background.

(2) Developing a data base adequate for use in model development and model validation is an extremely difficult task. It requires a very substantial investment in time, resources, and talent.

(3) Devising good ways to analyze and apply the data is at least as important as the data themselves. This often requires more advanced technical training than compilation of the data itself.

c. Probably the estimates of the amount of error in the data are overly confident (*i.e.*, the actual errors are larger than reported here). In this connection, see (among others) "The Reliability of Mortality Count and Suicide Count in the United States Army," William E. Dattel, *Military Medicine*, Vol 144, August 1979, pp 509-512; and "On Studying Morbidity Across All Three Armed Services," William E. Dattel, *Military Medicine*, Vol 146, August 1981, pp 590-590 (*i.e.*, one page in length).

d. There is a need for detailed quantitative data on strengths at intermediate stages during the course of a battle. Such information is not provided in the CDB90 data base.

e. Even the large CDB90 data base of battles is not large enough to support adequately all of the statistical analyses that should be performed.

f. At the time of this writing, CAA was in the process of developing, under contract with The Dupuy Institute (TDI), a similar data base for the southern portion of the World War II battle of Kursk. The initial phase of this effort, completed in January 1994, was directed at determining whether the requisite data could be obtained from the German and Russian archives. It concluded that the data could be obtained, and steps were being taken to move on to the second phase of actually preparing a completed data base.

1. **General Description.** Brinkerhoff, John R.; Dupuy, Trevor N.; Johnson, C. C.; and Bader, Brian R., "Casualty Estimates for Contingencies (CEC)," Data Memory Systems, Inc., Report No. CEC-0002AM, 15 November 1985, UNCLASSIFIED, in two volumes (41 pp and 211 pp). Report available from DTIC as AD-A183 255 (Vol 1) and AD-A183 256 (Vol 2). Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Includes data on personnel casualties in 81 contingency operations that took place from 1946 through 1982.

b. **Information Provided.** Information on each contingency operation includes the author's code number; contingency name; start and end dates, and duration in days; description of the type of contingency operation; distance advanced; identification of the two sides, together with their forces, insertion means, type of unit, and posture; descriptors for the terrain, weather, air superiority, and surprise; as well as strength and casualties. The casualty categories used are KIA, WIA, CMIA, other dead, other injured, and DNBI.

c. **Situational Descriptors Included.** In the original documentation, each contingency operation is described by a brief synopsis. The electronic form contains brief codes for the type of forces involved, the means of insertion, degree of opposition to the insertion, terrain, weather, air superiority, and surprise.

d. **Data Sources Used.** Various, but in most cases were secondary sources such as encyclopedias, dictionaries of battles, books on military history, newspaper reports, etc.

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 81 contingency operations and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Not used.

4. **Remarks.**

a. In many instances, data is recorded as unknown or not applicable, and in nearly all cases strengths or losses are uncertain to some degree, or have been estimated.

b. In some cases, the number of contingency operations reported may have been inflated by counting each phase or sector of an operation as a separate case. Also, the contingency operations reported were chosen because data on them was available. The user of these data must consider whether, or to what extent, these circumstances may have created a "selection bias."

COMPTON-1983

1. **General Description.** Compton, Jonathon, "Personnel Attrition Data," US Army TRADOC Systems Analysis Activity, White Sands Missile Range, New Mexico, Technical Report TRASANA, TR-29-82, ACN 67513, November 1983, 229 pp, UNCLASSIFIED-Limited, available from DTIC (AD-B085 206L). Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The diskette contains three files detailing personnel losses within the subunits of the US 6th Armored Division (Bastogne Defense, 31 December 1944 to 3 January 1945), the 7th Armored Division (St. Vith Defense, 17 to 23 December 1944), and the 28th Infantry Division (Schmidt Campaign, 2-13 November 1944). Information on personnel losses is developed down to company level within each division.

b. **Information Provided.** For each division and each day, the division's task organization and personnel losses are tabulated down to company level. Losses are identified as KIA, WIA, MIA, and nonbattle (NB)

c. **Situational Descriptors Included.** The tactical circumstances are recorded in a day-by-day narrative which includes the role that the four principal elements of tactical planning may have played in the battle: mission, enemy disposition, troops involved, and terrain (including weather).

d. **Data Sources Used.** Loss information was developed from corrected morning reports. Situational descriptors were obtained from various sources, but mostly from official US Army histories. For a detailed explanation of the sources used, see Compton, Jonathon, "Methodology for Collection of Personnel Attrition Data," Special Studies Division, US Army TRADOC Systems Analysis Activity, White Sands Missile Range, New Mexico, Special Studies Technical Memorandum TRASANA-TM-1-84, July 1984, 65 pp, UNCLASSIFIED, available from DTIC (AD-B122 270L).

e. **Diskette Format.** The diskette contains three files: COMP6AD.WQ! (Compton's 6th Armored Division data), COMP7AD.WQ! (Compton's 7th Armored Division data), and COMP28ID.WQ! (Compton's 28th Infantry Division data). Each file uses Compton's organization, i.e., daily, showing for each day the division's task organization and the personnel losses to each of its identified subunits.

3. **Other and Miscellaneous.** The source document includes numerous maps, organization charts, and a few other figures.

4. Remarks.

a. The organization of these files is not exactly the way a "data base" would be organized. Since the division's task organization (and, indeed, sometimes its basic composition) changed from time to time, it was necessary to show these changes. However, these dynamic changes precluded imposing a static data base organization on the data. Consequently, the computerized files merely reproduce the loss tables given in the source document. (Since the DTIC reproductions are always hard to read and sometimes illegible, these files make the data much more readily accessible.)

b. Compton provides no values for the number of personnel in the division as a whole, much less for those in a given subunit. Moreover, as noted in the preceding paragraph, the division's basic composition, as well as its task organization, changed from time to time as elements were attached and detached. Accordingly, casualty rates on a per person per day basis cannot be derived from these data.

c. Compton provides no data on the enemy's corresponding strengths or losses, even though these were the forces responsible for the friendly battle casualties.

d. In a few cases, Compton's values for losses do not add exactly to the totals he gives. However, the discrepancies are rare and in all cases minor (*i.e.*, generally amounting to a difference of only one or two).

e. Although Compton based his values on the corrected Morning Reports, in some cases he reports a large proportion of missing in action (MIA). Although their fate was not known at the time, surely some of these were actually KIA, others may have been WIA and then captured, others may have fallen unwounded into enemy hands, and some (wounded or not) may have returned to friendly control at a later date. Accordingly, the proper interpretation of these MIA data is unclear.

f. Compton was not able to locate all of the relevant data on the 28th Infantry Division. He notes (p 128 of the source document) that:

(1) "The following units were attached to the 28ID, but never brought into the battle area, and thus have not been included in the 28ID casualty tables:

502d Engineer Light Pontoon Company
993d Engineer Treadway Bridge Company
2d Platoon, 610th Engineer Light Equipment Company

(2) "Morning Reports for the 122th Infantry Regiment of the 12th Combat Team (attached to the 28ID for the period 3-10 November 1944) were not obtained by the author; thus, corresponding casualty data for the regiment is not included in the tabulations of losses."

CRETE-1993

1. **General Description.** Adapted from Engel, Joseph H., "Combat Effectiveness of Allied and German Troops in the World War II Invasion of Crete," Operations Evaluation Group Interim Research Memorandum IRM-35, 8 February 1963, UNCLASSIFIED, available from Center for Naval Analyses. Data base on diskette furnished to DTIC concurrently with this document.
2. **Data Base Description.**
 - a. **General.** Provides data on the German and Allied forces and losses during the Battle of Crete (D-day to D+13).
 - b. **Information Provided.** Information on each day includes the day (D+n); estimated number of German troops put ashore that day, together with their estimated losses and estimated strength at the end of the day; and corresponding information for the Allied side (except that the Allied "landings" were actually withdrawals). The casualty category used is losses.
 - c. **Situational Descriptors Included.** None, other than the day (D+n).
 - d. **Data Sources Used.** Various.
 - e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 92 battles and columns for each of the items of data provided.
3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the interim research memorandum cited in paragraph 1 above.
4. **Remarks.** The author warns that the estimated values may be quite inaccurate, since various sources report widely differing values. For other data bases purporting to give strengths and losses over time, see INCHON-1993, IWOJIMA-1993, POGOGORO-1993, and WESTWALL-1993.

CREWCAS-1991

1. **General Description.** Shephard, Ronald W., "Injuries to Tank Crews," Centre for Operational Research and Defence Analysis (CORDA Ltd), 233 High Holborn, London WC1V 7DJ, England. Report Number YR 4373. Work sponsored by the U.S. Army Model Improvement and Study Management Agency (MISMA), 1725 Jefferson Davis Highway, Suite 808, Crystal Square 2, Arlington, VA 22202. Report available from MISMA in the form of camera-ready copy, consisting of the following items:

Volume I, Western Desert 1941.

Volume II, Western Desert 1942, Part 1.

" " , Part 3.

" " , Part 4.

" " , Part 5.

" " , Part 6.

Volume III, Western Desert 1941-42, Part 1.

" " , Part 2.

" " , Part 3.

" " , Part 4.

" " , Part 5.

Volume IV, Burma 1945, Part 1.

Injuries to Tank Crews: A Guide to the Data Base and the Data Forms.

Diskette available from DTIC as AD-A265 499 (five diskettes, 3.5-inch, 730 KB, UNCLASSIFIED, March 1991).

2. Data Base Description.

a. **General.** This is a data base on casualties to over 850 tank crew personnel in World War II. The casualties were incurred from 183 cases in which tanks were hit or damaged in action (174 from the Western Desert during 1941-1942, and 9 from Burma during 1945). The data were extracted from official British military records. Each casualty is classified by injury severity; cause of damage; type of tank involved; position of the casualty; unit to which the tank belonged; disposition of the casualty; the weather, terrain, visibility, and location of the tank when hit; friendly and enemy tactics; consequences of the damage; and several other categories.

b. **Information Provided.** Detailed information on the type of tank hit, the unit to which it belonged, the theater of operation in which it was operating when hit, the range from which the shot was fired, the damage inflicted on the tank, the Shephard-Perkins index of injury severity for each crew member, type of ammunition used against the tank, the weather and terrain in the area, and a key to the source of the information provided. Time, terrain, and location of the tank when hit is reported if known. The casualty categories used are (i) killed; (ii) killed and wounded; (iii) killed, wounded, and missing; and (iv) losses. Other information is described in the documentation accompanying the diskettes.

d. **Situational Descriptors Included.** Time, terrain, location, visibility, and operational role of the tank when hit.

e. **Data Sources Used.** Various, but primarily official British military records and reports.

f. **Diskette Format.** The files are compatible with the dBASE III+ data base management system for IBM-PC compatibles operating under MS-DOS 3.3. They are provided on a set of five 3.5-inch, 730 KB diskettes. There are a total of 22 files on these diskettes, arranged as follows:

(1) Disk 1 contains:

- (a) CrewInj.dbf
- (b) CrewInj.dbt
- (c) TkCrV1p1.dbf
- (d) TkCrV1p1.dbt

(2) Disk 2 contains:

- (a) TkCrV2p2.dbf
- (b) TkCrV2p2.dbt
- (c) TkCrV2p3.dbf
- (d) TkCrV2p3.dbt

(3) Disk 3 contains:

- (a) TkCrV2p4.dbf
- (b) TkCrV2p4.dbt
- (c) TkCrV2p5.dbf
- (d) TkCrV2p5.dbt

(4) Disk 4 contains:

- (a) TkCrV3p2.dbf
- (b) TkCrV3p2.dbt
- (c) TkCrV3p3.dbf
- (d) TkCrV3p3.dbt

(5) Disk 5 contains:

- (a) TkCrV3p4.dbf
- (b) TkCrV3p4.dbt
- (c) TkCrV3p5.dbf
- (d) TkCrV3p5.dbt
- (e) TkCrV4p1.dbf
- (f) TkCrV4p1.dbt

g. **Using these Files to Set Up the dBASE III+ Files.** The files CrewInj.dbf and CrewInj.dbt are complete dBASE III+ data base files in themselves. However, the other files listed above (i.e., the files TkCrV?p?.dbf and TkCrV?p?.dbt) are not complete dBASE III+ data base files in themselves. Instead, they must be used to reconstitute the complete TkCrInjs.dbf and TkCrInjs.dbt dBASE III+ data base files. The procedure for this is described in the data base format documentation "Injuries to Tank Crews: A Guide to the Data-base and the Data Forms," prepared by Rowland Charles Goodman, Reference: YR 4373 Annex A, March 1991, provided by DTIC along with the data diskettes.

3. **Other and Miscellaneous.** None.

4. **Remarks.** We note that WALKER-1991 provides a data base of tank crew injuries from the Vietnam War.

DODGE-1993

1. **General Description.** Dodge, Theodore Ayrault, *Napoleon: A History of the Art of War from Lutzen to Waterloo, With a Detailed Account of the Napoleonic Wars*, (Vol 4 of four volumes), AMS Press, Inc., New York, NY 10003, 1970, reprinted from the edition of 1904-1907, UNCLASSIFIED, available from libraries. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The diskette contains one file in spreadsheet form detailing personnel strengths and losses as tabulated in the source's Appendix E.

b. **Information Provided.** The starting and ending dates and personnel strengths and losses are tabulated for the forces involved in battles from Breitenfeld (1631) to Waterloo (1815).

c. **Situational Descriptors Included.** Not tabulated, but discussed in the text portion of the source.

d. **Data Sources Used.** Various.

e. **Diskette Format.** The diskette contains one file in spreadsheet form, with rows for the forces reported by the source. There are 261 rows. The columns provide space for the following items of information: sequence number (an arbitrary number used to identify the row); the number of the page of the source from which the information is taken; the name of the battle; the battle's starting and ending dates; identification of the force; the personnel strength and casualties to that force; and the type of casualty reported (specifically, either killed or killed and wounded).

3. **Other and Miscellaneous.** Not used.

4. **Remarks.** The author's purpose in providing this information is to illustrate and support his claim that a force's percentage of casualties declines in a more or less regular fashion as the size of the force increases. Thus, the author proposes that the "usual" casualty percentage in "stoutly contested" battles is 20 percent for forces up to 5,000 men, 18 percent for forces of 5,000 to 10,000 men, 16 percent for forces of 10,000 to 20,000, 15 percent for forces of 20,000 to 30,000, 14 percent for forces of 30,000 to 40,000, 13 percent for forces of 40,000 to 60,000, 12 percent for forces of 60,000 to 100,000, 11 percent for forces of 100,000 to 130,000, and 10 percent for forces over 130,000 men. His appendix also cites some examples of particularly high loss percentages to moderately-sized units (3,000 to 10,000 men, and also to regimental or light brigade forces).

ICPSR-1984

1. **General Description.** "Wages of War, 1816-1980: Augmented With Disputes and Civil War Data," ICPSR data base number 9044, Inter-University Consortium for Political and Social Research (ICPSR), P.O. Box 1248, Ann Arbor, Michigan 48106, UNCLASSIFIED. US Department of Defense components can obtain a copy from DTIC (AD-M200 224, distribution authorized to DOD components only; proprietary data; January 1994. Other requests shall be referred to US Army Concepts Analysis Agency). Others can obtain copies from ICPSR.

2. **Data Base Description.**

a. **General.** The history of these data sets is summarized here. During the early 1960s, Professor J. David Singer initiated a series of studies at the University of Michigan that evolved into the Correlates of War project. This project was inspired by the quantitative analyses of war of such scholars as Quincy Wright, Lewis Richardson and Pitirim Sorokin. The project's aim is to identify the variables that have been most frequently associated with war since the Congress of Vienna in 1816. In 1974, the ICPSR released a collection of data sets by principal investigators J. David Singer and Melvin Small that was an outgrowth of the project. This collection was entitled "Wages of War, 1816-1965," and presented data on interstate wars.

The ICPSR data base number 9044 updates the earlier Wages of War collection, but in contrast to the original collection, this revised version presents separate data on civil wars and interstate disputes in addition to interstate wars (except that the data on international disputes is currently being withheld by ICPSR pending revision by the principal investigators).

The data were collected to study the trends and changes in the frequency, magnitude, severity, and intensity of interstate wars, civil wars, and interstate disputes. For each data set, the unit of analysis is the participant in a particular conflict. While the three data sets are related, they are mutually exclusive in that each describes a particular type of war (interstate or civil) or a dispute.

b. **Information Provided.** The data set on interstate wars (DA9044IS) describes the experience of each interstate system member in each war. The file contains 302 records. Each record contains the values for 42 variables that describe the experience of one nation in an interstate war.

The data set on civil wars (DA9044CW) is a study of 106 major civil wars involving 139 participants between 1816 and 1980. This file contains 139 records, each containing 36 variables.

A third data set describes the international disputes that occurred from 1817 through 1976. International disputes are distinguished from international wars in that fewer than 1,000 total fatalities are incurred by all the disputants involved. (This data set is currently being withheld by ICPSR pending revision by the principal investigators.)

c. Situational Descriptors Included.

(1) The data set DA9044IS for interstate wars includes the information regarding the experience of each interstate system member in each war from 1816 to 1980. The unit of analysis is the participant in a particular conflict. The file contains 302 records. Each record contains the values for 40 variables that describe the experience of one participant in an interstate war. In order to be considered a nation-participant in the interstate system, certain minimal criteria of population and diplomatic recognition are used. Qualifying nation-participants are classified as to whether or not they were members of the European central system at the time of the war and, therefore, active and influential in European-centered diplomacy. The geographical location of the war is coded as well as the severity of the war as determined by its duration and the number of battle-connected deaths. The prewar population of each participant is also coded.

Also included in this data set are two subtypes of wars in which a nation that qualifies as an interstate system member engages in a war with a political entity that is not an interstate system member. These are extra-systemic wars. The first extra-systemic war subtype, the imperial war, involves an adversary that is an independent political entity but does not qualify as a member of the interstate system because of serious limitations on its independence, a population that was not large enough to meet the interstate system membership criteria, or a failure of other states to recognize it as a legitimate member.

The second subtype, the colonial war, includes international wars in which the adversary was a colony, dependency, or protectorate composed of ethnically different people and located at some geographical distance or, at least, peripheral to the center of government of the given system member.

Both imperial and colonial wars are so designated in the list of interstate wars in the code book.

(2) The data set DA9044CW for civil wars includes the information regarding 106 major civil wars involving a total of 139 participants between 1816 and 1980. This file contains 139 records, each containing 36 variables. An internal war is classified as a major civil war if (a) military action was involved, (b) the national government at the time was actively involved, (c) effective resistance (as measured by the ratio of fatalities of the weaker to the stronger forces) occurred on both sides, and (d) at least 1,000 battle deaths resulted during the civil war.

The geographical area in which the war was fought is also coded as well as whether nations outside the civil war actively and overtly participated on one side or the other. The duration, beginning and ending dates of the civil war, and the prewar population and number in the armed forces of each participant are also included.

d. Data Sources Used. Various.

c. **Diskette Format.** The diskette contains five files as follows:

- (1) README.AAB, an ASCII file describing the diskette files.
- (2) CB9044IS.AAB, an ASCII file describing the file format and coding system for the data in file DA9044IS.
- (3) DA9044IS, an ASCII file containing the data on interstate wars.
- (4) CB9044CW.AAB, an ASCII file describing the file format and coding system for the data in file DA9044CW.
- (5) DA9044CW, an ASCII file containing the data on civil wars.

The code books (files CB9044IS.AAB and CB9044CW.AAB) occasionally refer to an OSIRIS format. However, because the files on the ICPSR-1984 diskette use only the card image format, also mentioned in the code books, all references to the OSIRIS format can be ignored.

The ICPSR-1984 card image files contain several decks per case in a format based on 80 column punched cards. However, the data are sorted by case with all decks for a case together. The deck number is given in columns 1 and 2. Thus, the card images are sorted in such a way that the numbers in the first two columns run from 01 to 03 (for the interstate war data set, DA9044IS) or from 01 to 05 (for the civil war data set, DA9044CW), and then repeat.

The next four columns, columns 3 through 6, give the ICPSR catalog number (9044, for both the interstate and civil war data sets). The entries in the other columns are as explained in the code book files, i.e., in the files whose names start with the letters "CB".

3. **Other and Miscellaneous.** Examples of the use of these data sets, as well as much important additional information regarding them, are contained in the following books:

a. Singer, J. David, and Small, Melvin, *The Wages of War: A Statistical Handbook*, John Wiley & Sons, Inc., New York, 1972, ISBN: 0-471-79300-0, 419 pp, UNCLASSIFIED, available from libraries.

b. Small, Melvin, and Singer, J. David, *Resort to Arms: International and Civil Wars, 1816-1980*, Sage Publications, Beverly Hills, CA, 1982, ISBN: 0-8039-1776-7, 372 pp, UNCLASSIFIED, available from publisher and libraries.

c. Singer, J. David (ed), *Correlates of War*, in two volumes, The Free Press (a division of Macmillan Publishing Co., NY); Vol I, 1979, ISBN: 0-02-928960-2, UNCLASSIFIED, available from libraries; Vol II, 1980, ISBN: 0-02-92010-4, UNCLASSIFIED, available from libraries (CAA Lib. No. = 327.11 COR).

4. Remarks.

a. By agreement between the ICPSR and CAA, distribution of the ICPSR-1984 data bases through DTIC is limited to DOD agencies only. Furthermore, DOD users are requested to provide the ICPSR with copies of (or at least with bibliographic citations to) any reports or conference proceedings prepared using these data sets. Non-DOD agencies may obtain copies directly from ICPSR upon payment of the appropriate fees.

b. When supplemented with additional information gleaned from the books cited in paragraph 3 above, this appears to be the best currently available data base on interstate and civil wars.

INCHON-1993

1. **General Description.** Busse, James J. "An Attempt to Verify Lanchester's Equations," Paper 9-1 (pp 587-597) in Volume 2 of *Developments in Operations Research*, Benjamin Avi-Itzhak (Executive Editor), Gordon and Breach Science Publishers, New York, 1971 (ISBN 0-677-30840-X), UNCLASSIFIED, available from libraries. This is based on an earlier paper bearing the same title and prepared while Busse was on the staff of the Marine Corps Operations Analysis Group of the Center for Naval Analyses. This earlier paper is undated, but appears to have been written sometime around 1968. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The data base provides Busse's values for the US Marine Corps strength at the end of each day of the campaign (14 September 1950 to 3 October 1950); the estimated number of North Korean reinforcements arriving during each day; and the estimated North Korean strength at the end of each day.

b. **Information Provided.** As described in paragraph 2a above. The casualty category used is losses.

c. **Situational Descriptors Included.** None, other than the day of the campaign.

d. **Data Sources Used.** The data appear to have been obtained from *US Marine Operations in Korea 1950-1953*, Volume II: "The Inchon-Seoul Operation," Lynn Montross and Captain Nicholas A. Canzona (USMC), Historical Branch, G-3, Headquarters, US Marine Corps, Washington, DC, 1955. The information used appears to stem specifically from Appendix B (Build-up of 1st Marine Division (Reinf), Appendix C (Task Organization of Marine Division for Inchon Landing), Appendix F (Final Troop List of Division for the Inchon Landing), Appendix H (Enemy Units During the Inchon-Seoul Campaign), and Appendix J (Casualties During the Inchon-Seoul Campaign).

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the days and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the papers cited in paragraph 1 above and in Hartley, Dean S. III, "Can the Square Law be Validated?," Data Systems Research & Development Program, Oak Ridge Gaseous Diffusion Plant, operated by Martin Marietta Energy Systems, Inc. for the US Department of Energy, K/DSRD-57, March 1989, 72 pp, UNCLASSIFIED, available from the publisher and DTIC.

4. Remarks.

a. The data reported by and used by Busse, and subsequently by others, while lifted from the source cited in paragraph 2e above, or some equivalent source, seems to be somewhat at variance with the historical facts. Some of the problems with Busse's data are detailed in the following subparagraphs.

b. Busse's figure of 25,040 for the initial strength of the US Marines is taken from the total given for the final troop list of Division for the Inchon landing. There are three caveats to this figure, as follows.

(1) In the first place the source itself describes its tabulation as follows (emphasis added): "Below, as nearly as can be determined, is the troop list of the 1st Marine Division (Reinf) for the Inchon Landing. The list is correct as to units included, *but has been difficult to reconcile the totals given for units.*" (Emphasis added.)

(2) Secondly, it is not credible that the whole force of 25,040 was landed on D-day (15 September 1950). The figure of 15,279 total forces ashore at the end of D-Day [given in Commander in Chief, US Pacific Fleet Interim Evaluation Report No. 1, Korean War, US Pacific Fleet Operations. Period 25 June to 15 November 1950 (Vol 4: Marine Air; Vol 5: Amphibious and Ground)] is much more realistic.

(3) Busse's figure of 25,040 for the Marine division's strength includes its Korean Marine Corps Regiment (2,786 men), the attached US Army units (2,760 men), and a few US Navy units. Thus, some 5,546 men (or about 22 percent of the division's strength), although associated with the division, were in other than USMC units. Unfortunately, it is not clear whether the casualties reported include those taken by non-USMC units.

c. Nor is Busse's figure of 22,150 for the North Korean strength at the start of D-day realistic—instead it is clear that hardly more than 2,200 enemy troops were in the immediate vicinity of Inchon at that time (counting troops in Inchon proper and those garrisoning the island, known as Wolmi-do, guarding Inchon harbor).

d. Other United Nations units were also landed and participated in the battle, although their presence and influence on the action is not accounted for anywhere in Busse's figures. These units included at least the combat units listed below. The source cited in paragraph 4b(2) above states that a total of 79,100 personnel were landed at Inchon between 16 September and 5 October (D-day to D+20), although many of these would have been support units. Robert Debs Heinl, Jr.'s book *Victory at High Tide: The Inchon-Seoul Campaign*, The Nautical and Aviation Publishing Co. of America, 3d Edition, 1979, states that there were 53,882 "troops on the beach" as of 22 September (D+7).

(1) The US Army 7th Infantry Division began coming ashore on 17 September 1950 (D+2). According to Table 22 (pg 23) of Reister's book [Frank A. Reister, *Battle Casualties and Medical Statistics: U.S. Army Experience in the Korean War*, The Surgeon General, US Army, 1986] this division, in 13 days of combat from 18 to 30 September (D+3 to D+15), suffered 94 KIA and 319 WIA.

(2) The US Marine Corps Regimental Combat Team RCT-7, part of the 1st Marine Division, landed 21 September (D+6).

(3) The US Army 17th Infantry Regiment landed somewhat later than 21 September (D+6).

(4) The US Army 187th Airborne Regimental Combat Team was flown into Kimpo Airfield on 24 and 27 September (D+9 and D+12).

(5) Linkup between the Inchon-Seoul beachhead and the US Eighth Army (from the Pusan Perimeter) occurred on 26 September (D+11).

e. Busse's schedule of North Korean reinforcements does not agree with that given in the source cited in paragraph 2d. In particular, Busse lists North Korean units as opposing the UN forces as soon as they are "present" or "formed," but before their presence is "identified" by the taking of prisoners and hence before they were engaged. However, the source itself is in many respects ambiguous and is plainly inconsistent in at least one instance, in that it refers a North Korean 18th Rifle Division as being "identified" on 16 September (D+1) but locating it in the Seoul area, the extreme western regions of which were not approached until 22 September (D+7). This source also mentions a North Korean 17th Rifle Division as being first identified on 26 September, but does not include such a unit in its Order of Battle for the enemy units (it may be that this is a misprint for the 18th Rifle Division).

f. For other data bases purporting to give strengths and losses over time during a battle, see CRETE-1993, IWOJIMA-1993, POGOGORO-1993, and WESTWALL-1993.

IWOJIMA-1993

1. **General Description.** Helmbold, Robert L., "Iwo Jima Spreadsheet," US Army Concepts Analysis Agency, November 1993, UNCLASSIFIED, unpublished personal notes. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The diskette is based on one version of the landing schedule and remaining effective personnel strength of the US Marine Corps and Japanese units during the principal part of the battle of Iwo Jima. Other versions are possible, as explained in subsequent paragraphs.

b. **Information Provided.** The information covers the period from 19 February 1944 (D-day) to 16 March 1944 (D + 25), when the island was declared captured and occupied, and the operation was declared completed. For each day in this period, the data base gives a version of the estimated number of US Marine Corps troops put ashore, and a version of the estimated residual number of US Marine Corps troops (*i.e.*, those that had not become a casualty and were on the island). The estimated Japanese personnel strength at the start and end of this period are also given. US Marine Corps strengths, fitted to these data using a discrete approximation to the Lanchester square law, are also shown.

c. **Situational Descriptors Included.** Not used. See various narrative historical accounts of the operation for the tactical context.

d. **Data Sources Used.** Various data sources were used. The principal ones are listed below.

(1) The US Marine Corps landing schedule values were taken from Samz, Robert W., "Some Comments on Engel's 'A Verification of Lanchester's Law,'" *Operations Research*, Vol 20, No 1, January-February 1972, pp 49-52.

(2) The US Marine Corps personnel casualties are given as an average per day, based on information from Bartley, Whitman S., *Iwo Jima: Amphibious Epic*, Historical Branch, G-3 Division, Headquarters, US Marine Corps, 1954. Those for D-day (19 February 1944) are on p 68; for the period D + 1 to D + 5 (20-24 February 1944) on p 98; for D + 6 through D + 19 (25 February to 10 March 1944) on pp 122, 148, and 177; for the period D + 20 through D + 35 (11-26 March 1944) on p 193.

(3) The Japanese initial personnel strength adopts the figure cited by Morehouse, Clifford P., *The Iwo Jima Operation*, Historical Division, Headquarters, US Marine Corps, 1946 (specifically on p 146).

(4) The Japanese personnel strength at the end of D + 25 (17 March 1944) is based on Bartley's statement (p 193) that, despite the island's having been declared captured on that date, "In April and May, however, aggressive patrol and ambush activity by the 147th Infantry [US Army] netted 867 prisoners and 1,602 Japanese killed."

c. **Diskette Format.** The diskette contains one spreadsheet with rows for the days and columns for the other information provided.

3. **Other and Miscellaneous.** Several other versions of the strengths and losses at Iwo Jima can be constructed from the available sources. Some of the relevant considerations are mentioned below.

a. **US Marine Corps Landing Schedule.** Engel, Joseph H., "A Verification of Lanchester's Law," Operations Research, Vol 2, 1954, pp 163-171, gives the following landing schedule for Iwo Jima:

<u>On Day</u>	<u>Troops Landed</u>
D	54,000
D + 1	0
D + 2	6,000
D + 3	0
D + 4	0
D + 5	13,000
<u>all other</u>	<u>0</u>
Total	73,000

However, Samz argues that the following schedule is far more likely to be closer to the truth:

<u>On Day</u>	<u>Troops Landed</u>
D	30,000
D + 1	1,200
D + 2	6,735
D + 3	3,626
D + 4	5,158
D + 5	13,227
D + 6	3,054
D + 7	3,359
D + 8	3,180
D + 9	1,454
D + 10	252
<u>Thereafter</u>	<u>0</u>
Total	71,245

Samz's landing schedule seems more supportable by the data. Morehouse (p 154A) reports that the landing force totaled 71,245. Bartley (p 67, footnote) remarks that 30,000 troops were put ashore on D-day. Morehouse (p 137) states that "The total Landing Force, including both Navy and Marine personnel, comprised about 70,000 men, of which 61,000 were combat Marines. Of these, 30,000 were landed on D-day, and all assault forces were ashore by D+5." For comparison, Samz's schedule puts a total of 59,946 troops ashore by the end of D + 5, which if interpreted as being all assault troops would agree reasonably well with Morehouse's statement.

On the other hand, Morison, Samuel Eliot, *History of United States Naval Operations in World War II*, Volume XIV, *Victory in the Pacific*, Little, Brown and Company, Boston, 1960, (p 72) makes the remark that "During the five weeks of the campaign, 7,500 battle replacements were provided, most of them coming under fire for the first time." Morehouse (p 86) also mentions the need for training of replacements by the 5th Marine Division "as the majority of them were not prepared for the type of fighting employed in this operation." The source of these replacements is not clear. The above landing schedules obviously presume that they were part of the Landing Force, rather than arriving after D + 10. Nevertheless, it is possible that they were drawn instead from Marine units held in reserve aboard ships in the vicinity and landed after D + 10. If so, the landing schedule should be revised to reflect these replacements.

Morehouse (p 87) states that on 18 March 1945 (D + 27) the 4th Marine Division, relieved by the 3d Marine Division, embarked on their ships and closed their command post, thus ending their participation in the battle. The same source reports that US Marine Corps artillery was secured and re-embarked "as rapidly as possible," with corps artillery completing embarkation on 17 March (D + 26) and the division artillery reembarked "on subsequent days." The timing of these reembarkations is the reason why the end of the day of 16 March 1945 (D + 25), or (what is the same thing) the beginning of D + 26, is used as the endpoint of the battle for comparisons of historical and fitted strengths.

b. **US Marine Corps Personnel Casualties.** The Bartley values, mentioned in the paragraph on data sources used, are as follows. Here and in all of our statements about Iwo Jima, unless explicitly stated otherwise, we include KIA, WIA, DOW, MIA, and combat fatigue.

<u>Period</u>	<u>Casualties</u>
D-day	2,420
D + 1 to D + 5	5,338
D + 6 to D + 19	11,930
<u>D + 20 to D + 35</u>	<u>3,885</u>
Total	23,573

Morison (p 69) gives figures amounting to 25,851 through D + 36. Bartley (pp 220 and 221) gives for the entire operation 22,056 for the Marine Corps casualties and 2,706 for the US Navy, for a total of 24,854. Morehouse (p 165D) gives a "grand total" of Landing Force casualties (KIA, WIA, MIA, and "ineffectives") of 21,082. However, on p 166D he gives a total of 20,860 for the KIA, WIA and MIA through D + 35. Engel uses a figure of 20,860 friendly casualties, but also used D + 28 (19 March 1945) for the end of the battle.

US casualties through D + 25 were estimated to be 21,145, arrived at by prorating Bartley's 3,885 casualties for the last period. The figure on Morehouse's p 166D for the same period is 19,928, which, however, does not include the combat fatigue cases. Since, according to the figures in Bartley, these

total 1,986 for the entire battle, there is a reasonable concordance between Bartley's and Morehouse's figures.

The US casualties given in the data base presume that they were uniformly distributed over the number of days in each of the respective time periods, rather than being bunched in any give day. In essence, this means that US casualties are presumed to have occurred at piecewise constant rates. This is surely not strictly correct, but is perhaps acceptable as a first approximation.

c. **Initial Japanese Personnel Strength.** Bartley (p 231) gives the range 20,350 - 21,060. Morehouse (p 146) estimates it at 21,449. The value implied by Morison's figures (p 69) amounts to 23,388. Engel (p 168) uses 21,500, as does Samz.

d. **Japanese Strength at the End of D + 25.** Bartley (p 193) states that, "Just what percentage of the Japanese defense force had become casualties by this time is not known, but as of 26 March [D + 35] only 216 prisoners of war had been taken. In April and May, however, aggressive patrol and ambush activity by the 117th Infantry [US Army] netted 867 prisoners and 1,602 Japanese killed." Morehouse (p 146) notes that of the 217 (sic) prisoners, 159 were Japanese and 58 Koreans—of this total, 11 died and an estimated 85 percent were wounded and required medical attention. These prisoners are counted as among the Japanese casualties incurred through D + 25. Morison (p 69) repeats Bartley's figures for the number of Japanese taken in April and May. Thus, at least $867 + 1,602 = 2,469$ Japanese were still alive and uncaptured at the end of the day 16 March 1945 (end of D + 25, or start of D + 26).

e. **Other Versions.** Various versions of the Iwo Jima casualty experience have been used in such studies as:

(1) Hartley, Dean S. III, "Confirming the Lanchestrian Linear-Logarithmic Model of Attrition," Data Systems Research & Development Program, Oak Ridge Gaseous Diffusion Plant, operated by Martin Marietta Energy Systems, Inc. for the US Department of Energy, K/DSRD-263/R1, February 1991, 136 pp, UNCLASSIFIED, available from DTIC. This is a revision to the report of the same title published in December 1990.

(2) Helmbold, Robert L., "Direct and Inverse Solution of the Lanchester Square Law With General Reinforcement Schedules," European Journal of Operations Research, to appear, pagination to be determined, UNCLASSIFIED, available from publisher and libraries.

(3) Helmbold, Robert L., "The Constant Fallacy: A Persistent Logical Flaw in Applications of Lanchester's Equations," European Journal of Operations Research, to appear, 1993, pagination to be determined, UNCLASSIFIED, available from publisher and libraries.

4. **Remarks.** In view of the conflicting reports, it is difficult to establish exact figures for Iwo Jima. While perhaps as good as any, the ones proposed here are but one possible version, and many others are also possible and perhaps equally plausible. For many years, Iwo Jima was the only known example of a battle in which estimated strengths and losses to both sides were available in enough detail to be used to provide a valid test the form of attrition law. The data have been used on many occasions for such purposes. For other data bases purporting to provide data on strengths and losses over time, see CRETE-1993, INCHON-1993, POGOGORO-1993, and WESTWALL-1993.

KELLEY-1977

1. **General Description.** Kelley, C. T.; Smith, G. K.; and Matyskiela, W. W., "Quality-Quantity Tradeoffs: A Historical Analysis of Air Combat," The RAND Corporation Report R-1976-AF, April 1977, 45 pp, UNCLASSIFIED. Report available from DTIC as AD-C010 504. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Contains data for both sides in 315 air combat events from the Korean War that occurred between 19 August 1951 to 22 July 1953, and on 138 air combat events that occurred during the Vietnam War.

b. **Information Provided.** For the Korean War, the data include the date on which the event took place together with the numbers and losses to both sides in each air combat event. For the Vietnam War, the date of the event is not given, but the data include an identification of the type of aircraft used by each side, as well as their numbers and losses. The Project Red Baron event number is given for the Vietnam War data. The casualty category used is aircraft losses.

d. **Situational Descriptors Included.** None.

e. **Data Sources Used.** The Vietnam War data are based on "Project Red Baron—Air-to-Air Encounters in Southeast Asia," USAF Tactical Fighter Weapons Center. The Korean War data are based on Fifth Air Force Intelligence Summaries.

f. **Diskette Format.** The diskette contains one file in spreadsheet form with one row for each of the air combat engagements and columns for each of the items of data provided on them.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the RAND report cited in paragraph 1.

4. **Remarks.** The data often show no aircraft losses to one or both sides as a result of the encounter. Whether any aircraft were damaged (but returned to base) is not recorded in these data. For the most part, the number of aircraft involved is relatively small (often either two, four, or eight aircraft), so that analyzing the data using differential attrition equations (as is attempted in the RAND report cited in paragraph 1) may be inappropriate—probabilistic models appear to be better suited to modeling this kind of data. Moreover, the analysis in the RAND report obtains the wrong results because it commits the Constant Fallacy

KRIVO-1993

1. **General Description.** Krivosheyev, G. F. (general editor), *Losses to the Armed Forces of the USSR in Battles, Combat Operations, and Military Conflicts: Statistical Investigations*, Military Press, Moscow, 1993 (in Russian), UNCLASSIFIED (formerly SECRET), 415 pp, book available from publisher and libraries. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The diskette contains information on losses in 51 major strategic operations conducted by the Soviet Union during World War II. The source document contains a great deal of additional information about Soviet losses in military actions from the Russian Revolutionary period to the end of the Soviet Union.

b. **Information Provided.** Losses are tabulated for 180 major Soviet military unit actions (primarily by Fronts) that occurred during these Soviet strategic operations. Each entry includes the page of the source document from which the data were taken, the type (offensive, defensive, or battle) and name of the major strategic operation, an identification of the major unit and the starting and ending dates of its participation in the major strategic operation, the unit's starting personnel strength, and the unit's permanent and medical losses. The casualty categories used are permanent and medical losses. Here the Soviet definitions of permanent and medical losses are as follows:

(1) **Permanent Losses.** These include personnel killed on the field of battle, died of wounds during medical evacuation, missing in action, or taken captive (the number of died of wounds or sickness in hospitals appears only in the total losses of the military forces). In the number of permanent losses to the army in the field are included also noncombat losses, not connected with immediately obvious combat operations—those who perish as a result of serious accidents, sentenced by military tribunals to capital punishment for various crimes against the state, committed suicide, or died of sickness in hospitals.

(2) **Medical Losses.** These include personnel wounded, contused, burned, sick, and frost-bitten in combat, who were evacuated from the zone of combat operations to Army, Front, or communications zone hospitals. Our total number of medical losses necessarily includes many soldiers who received two or more wounds, and were repeatedly given hospital treatment, each such treatment being counted as a separate loss.

c. **Situational Descriptors Included.** None, other than the name and dates of the engagement. However, the source document contains a brief synopsis of each major strategic operation and its results, including a statement of its duration in days, width of front, and depth of advance or retreat.

d. **Data Sources Used.** Various, but much use was made of the official Soviet military and state archives.

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 180 major units actions and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Not used.

4. **Remarks.** The source document contains many additional statistical tabulations of Soviet losses which were too voluminous to be reduced to digital form within the scope of this effort.

LIVERMORE-1993

1. **General Description.** Adapted from Livermore, Thomas L., *Numbers and Losses in the Civil War in America, 1861-65*, Kraus Reprint Co., Milwood, NY, 1977 (a reprint of the Indiana University Press edition of 1957, originally published in Boston in 1900), 150 pp, UNCLASSIFIED, available from publisher and libraries. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Provides data on strengths and losses in 63 battles of the American Civil War, from Bull Run (21 July 1861) to Petersburg II (2 April 1865).

b. **Information Provided.** Information on each battle includes the battle name; its starting and ending dates; identification of which side was the attacker; the personnel strengths, KIA, WIA, and CMIA; and an identification of which side won. The casualty categories used are KIA, WIA, and CMIA.

c. **Situational Descriptors Included.** None, other than the name and dates of the engagement.

d. **Data Sources Used.** Based on the official records of the War of the Rebellion.

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 63 battles and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group, Staff Paper CORG-SP-128, 1 July 1961, 178 pp, UNCLASSIFIED. Report available from DTIC as AD-480 975.

4. **Remarks.** The source has a reputation for being one of the most careful and detailed assessments of strengths and losses of battles in the American Civil War.

LMI-1989

1. **General Description.** Kuhn, George W. S., "Ground Forces Battle Casualty Rate Patterns: The Empirical Evidence," Logistics Management Institute Report FP703TR1, September 1989, 270 pp, UNCLASSIFIED, report available from Logistics Management Institute (LMI), 2000 Corporate Ridge Parkway, Tysons Corner, VA 22102. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** This data base records the US personnel strengths and battle casualties for operations in Northwest Europe during World War II. The data include both the US 1st Army (when it was part of GEN Montgomery's 21st Army Group prior to the Normandy breakout) and the US 12th Army Group (commanded by General Omar Bradley and formed in August 1944). The dates covered are from 15 June 1944 through 30 April 1945 (although no data are available for 16 June 1944, 3 July 1944, 6 July 1944, and 11 August 1944 through 30 September 1944). Personnel strength and loss data are provided for each division and each major higher echelon unit (such as corps, army, or army group headquarters), as well as for the headquarters/service elements and nondivisional combat units assigned/attached directly to those commands. There are a total of 14,552 records in the file. An analyst may focus on each division individually. Alternatively, by using a data base manipulation program such as dBase^{III}, the analyst can create gain numerous perspectives on the data (for example, by creating daily records for corps and higher echelons).

b. **Information Provided.** Each daily record includes the date; the type of element involved (infantry, armor, airborne, headquarters/service, or other); numerical designation of the division, corps, army, or army group to which those elements belonged; the number of KIA, WIA, CMIA for this those elements; and the authorized and assigned personnel strength of those elements as of the date of this record. The casualty categories used are KIA, WIA, and CMIA.

c. **Situational Descriptors Included.** None, other than the date.

d. **Data Sources Used.** Division strength and casualty data as reported in contemporary G-1 Daily Summaries, obtained from US archival materials. See the report cited in paragraph 1 (especially Chapter 5 and Appendix B) for a full discussion of the sources used.

e. **Diskette Format**

(1) The diskettes made available to CAA contain the following files, which were originally designed for use with the dBaseTM data base system: 12ARGP.TXT and ALT_RES.TXT. The records in the 12ARGP.TXT files have the following field structure (the example is purely hypothetical):

<u>Field</u>	<u>Field Name</u>	<u>Type</u>	<u>Width</u>	<u>Example</u>
1	DATE	Date	8	19441220
2	UNIT	Character	5	INF D
3	DIV	Numeric	3	99
4	CORP	Numeric	2	5
5	ARMY	Numeric	2	1
6	ACP	Numeric	2	12
7	KIA	Numeric	6	15
8	WIA	Numeric	6	65
9	CMIA	Numeric	6	303
10	AUTH	Numeric	6	14184
11	ASSD	Numeric	6	10734
12	VERSION	<u>Character</u>	<u>2</u>	
		Total	54	

Here the DATE is given in the form YYYYmmdd. UNIT denotes the type of basic unit. For divisions, UNIT designations are either INF D, ARM D, or ABN D. For higher echelons, the UNIT designations are either INF D, ARM D, or ABN D (for individual divisions attached directly to the higher echelon); HQ/SE (for organic and attached "headquarters and service" elements); or OTHER (for other *combat* elements organic or attached to the higher echelon, such as artillery, tank destroyer units, *etc.*—although some of these were often temporarily assigned or attached to divisions for operational control, this data base carries them at the higher echelon). Numerical designations are used for the DIV, CORP, ARMY, and AGP (Army Group) fields. The number of KIA, WIA, and CMIA for the indicated DATE are provided. The number of personnel AUTH (Authorized) and ASSD (Assigned) to the unit on the indicated date are given. Missing numerical values are coded as -99.

The VERSION is a special code used to distinguish between three alternative versions of the data for the 106th Infantry Division for the period 12-25 December 1944, and two alternative versions of the 7th Armored Division during the period 21-24 December 1944. The first character of this code (column 53) indicates who developed the version:

R, value reported in the 12th Army Group's G-1 Daily Summaries,

L, value developed by LMI,

D, value developed by DMSI (*see* the ACSDB-1990 data base),

T, value developed by TRASANA (*see* the COMPTON-1983 data base).

The second character of the version code (column 54) indicates which division the version applies to: 6 for the 106th Infantry Division, or 7 for the 7th Armored Division. For example, the versions provided in the main file are those used by LMI, and are coded as L6 (LMI version for the 106th Infantry Division), and R7 (taken from the 12th Army Group's archival records for the 7th Armored Division). The other versions are provided in the ALT_RES.TXT file, and are coded as D6 (DMSI's version for the 106th Infantry Division), R6 (taken from the 12th Army Group's records for the 106th Infantry Division), and T7 (Compton's TRASANA study version for the 7th Armored Division).

It is known that division and higher commands did not record the surrender of major elements of the 106th Infantry Division, which occurred on 19 December 1944, until 24 December 1944. Hence, the R6 version does not correspond to reality for the period 19-23 December 1944. Compton's TRASANA study provides a detailed compilation of strengths and losses of the 7th Armored Division, based on the Morning Reports made by its subordinate company-level units. Each of the alternative versions is given below.

Date	Unit	No	Corps	Army	ArGrp	KIA	WIA	CMIA	Auth	Assd	Code
19441217	INF D	6	8	1	12	52	105	290	14184	13343	D6
19441218	INF D	6	8	1	12	38	105	155	14184	12926	D6
19441219	INF D	6	8	1	12	661	104	6390	14184	5650	D6
19441220	INF D	6	8	1	12	37	0	370	14184	5243	D6
19441221	INF D	6	18	1	12	14	94	10	14184	5104	D6
19441222	INF D	6	18	1	12	1	12	3	14184	5056	D6
19441223	INF D	6	18	1	12	4	20	13	14184	4995	D6
19441224	INF D	6	18	1	12	2	16	0	14184	4965	D6
19441225	INF D	6	18	1	12	0	7	0	14184	4944	D6
19441217	INF D 106	8	1	12	72	390	1356	14184	12180	L6	
19441218	INF D 106	8	1	12	83	289	406	14184	12407	L6	
19441219	INF D 106	8	1	12	203	403	6549	14184	5252	L6	
19441220	INF D 106	8	1	12	14	66	215	14184	4957	L6	
19441221	INF D 106	18	1	12	14	94	10	14184	4839	L6	
19441222	INF D 106	18	1	12	1	12	9	14184	4817	L6	
19441223	INF D 106	18	1	12	4	20	13	14184	4780	L6	
19441224	INF D 106	18	1	12	2	16	0	14184	4586	L6	
19441225	INF D 106	18	1	12	0	7	0	14184	5067	L6	
19441217	INF D	6	8	1	12	72	390	1356	14184	12180	R6
19441218	INF D	6	8	1	12	83	289	406	14184	12407	R6
19441219	INF D	6	8	1	12	74	210	646	14184	11682	R6
19441220	INF D	6	8	1	12	14	66	215	14184	10337	R6
19441221	INF D	6	18	1	12	14	94	10	14184	10895	R6
19441222	INF D	6	18	1	12	1	12	3	14184	10895	R6
19441223	INF D	6	18	1	12	4	20	13	14184	10895	R6
19441224	INF D	6	18	1	12	2	16	0	14184	4586	R6
19441225	INF D	6	18	1	12	0	7	0	14184	5067	R6
19441221	ARM D	7	18	1	12	3	25	0	10729	10542	R7
19441222	ARM D	7	18	1	12	0	10	0	10729	10542	R7
19441223	ARM D	7	18	1	12	18	90	80	10729	10542	R7
19441224	ARM D	7	18	1	12	12	56	124	13822	13401	R7
19441221	ARM D	7	18	1	12	7	19	200	10729	10384	T7
19441222	ARM D	7	18	1	12	3	68	26	10729	10362	T7
19441223	ARM D	7	18	1	12	2	41	92	10729	10160	T7
19441224	ARM D	7	18	1	12	12	56	124	13822	9967	T7

(2) According to Eisenhower's *Crusade in Europe*, by the end of the war, GEN Bradley's 12th Army Group was organized as shown below. However, by that time its organization had changed considerably since the days of the Normandy landings.

12th Army Group (Bradley)
 3d Army (Patton)
 8 Corps (Middleton)
 12 Corps (Eddy)
 20 Corps (Walker)
 1st Army (Hodges)
 3 Corps (Van Fleet)
 5 Corps (Huebner)
 7 Corps (Collins)
 18 Abn Corps (Ridgway)
 15th Army (Gerow)
 22 Corps (Harmon)
 23 Corps (Balmer)
 9th Army (Simpson)
 13 Corps (Gillem)
 16 Corps (Anderson)
 19 Corps (McLain)

The data base can be used to track the daily changes in structure of the US forces, including the changing army assignments to army groups, changing corps assignments to armies, and changing division assignments, including their arrival in theater and subsequent reassignments. Thus, it begins on 15 June 1944 with the 1st US Army and its 5th, 7th, 8th, and 19th Corps in Normandy when the 1st US Army was under GEN Montgomery's 21st Army Group, and (except for the missing dates noted above in paragraph 2a) continues through the formation of the US 12th Army Group until 30 April 1945.

(3) Apart from the special DIV code 0 (nondivisional), the divisions listed in the *.TXT files made available to CAA are:

Airborne Divisions: 17, 82, 101.
 Armored Divisions: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 20.
 Infantry Divisions: 2, 4, 5, 8, 9, 26, 28, 29, 30, 35, 44, 65, 66, 69, 70,
 71, 75, 76, 78, 79, 84, 86, 87, 89, 90, 94, 95, 97, 99, 102, 104, and 106.

(4) The echelon above division for which strengths and losses of their "HQ/SE" and "OTHER" elements are listed in the *.TXT files made available to CAA are as follows:

Corps: 3, 5, 7, 8, 12, 13, 16, 18, 19, 20, 22, and 23.
 Army: 1, 3, 9, and 15.
 Army Group: 21 listed as 21 when, prior to formation of the 12th Army Group, US forces operated under the British 21st Army Group).

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following studies:

- a. "Ground Forces Battle Casualty Rate Patterns: The Empirical Evidence," Logistics Management Institute Report FP703TR1, September 1989.
- b. "Ground Forces Battle Casualty Rate Patterns: Current Rate Projections Compared to Empirical Evidence," Logistics Management Institute Report FP703TR2, May 1990.
- c. "Ground Forces Battle Casualty Rate Patterns: Suggested Planning Considerations," Logistics Management Institute Report FP703TR3, January 1991.
- d. "Ground Forces Battle Casualty Rate Patterns: Uses in Casualty Estimation and Simulation Evaluation," *Proceedings of the 1992 60th Military Operations Research Symposium*, to be published.
- e. "Capability-related Medical Support Planning," SHAPE Headquarters, 2105.05/SHOLM/055/93, 9 March 1993.

4. **Remarks.**

a. This data base is unique in that it facilitates viewing and comparing casualty rate experience at both the tactical (division and corps) and operational (army or army group) levels of combat on both a daily basis and over more extended periods of time. The use of a suitable data base system facilitates viewing losses in any of several modes. For example, losses can be viewed for single divisions, for divisions of certain types, for divisions grouped under one or more of three higher echelons (corps, army, and army group), for the higher echelons considered apart from their divisional personnel (looking either at their HQ/service troops alone, at their nondivisional combat troops alone, or at the combination of these two categories), or for an echelon "slice" (by combining the two echelon-unique categories with its assigned lower echelon units). However, the data are purely one-sided (i.e., they do not record what was happening on the opposing side). The data provided overlap to some extent that provided in the ACSDB, and it would be of some interest to compare the two in detail to see the extent to which their strengths and losses agree (for a comparison of the losses only, see pp B-75 to B-78 of the source document).

b. Every data base has some special characteristics of which the user should be aware. Some of the special characteristics of this data base are mentioned here. First, there are the gaps in the available source data, mentioned in paragraph 2a, above. Second, prior to the formation of the US 12th Army Group, the Army Group number is properly given as 21, a British Army Group. Third, there are no HQ/SE elements listed in the data base for the period prior to 1 October 1944—although some US corps and Army HQ/SE must have been active prior to that date, the sources used to prepare this data base provide no information on them. Fourth, the 44th Infantry Division appears only briefly (in 3 Corps of the 9th Army from 1 October 1944 through 10 October 1944 and as attached directly to the

9th Army from 11 October 1944 through 16 October 1944). Throughout its appearance, all of its battle casualties are recorded as zero, and its assigned strength varies from a low of 13,551 to 13,895. This agrees with the information in the sources consulted in preparing the data base. Fifth, for the period 23 January 1945 through 31 January 1945, the 35th Infantry Division in 3 Corps of the 3d Army consisted of a regimental combat team only, which explains its unusually low authorized and assigned strength values. Sixth, at various times from 2 October 1944 through 24 October 1944, the 9th Armored Division in 3 Corps of the 9th Army (later reassigned to 8 Corps of the 3d Army) had one of its combat commands assigned to another division, which explains its occasional unusually low authorized and assigned strength values. Seventh, at various times from 15 June 1944 through 11 July 1944, both the 82d and 101st Airborne Divisions had unusually low assigned strengths, but these have been confirmed by a check against the original source materials used. Eighth, at various times from 19 December 1944 through 6 April 1945, the 106th Infantry Division had unusually low assigned strengths, but these were a direct result of the heavy losses this division took in the early phases of the Ardennes Campaign (Battle of the Bulge). Apparently, this division was never restored to full strength after suffering these losses and was either offline or assigned to a quiet sector thereafter.

MOTTELAY-1993

1. **General Description.** Mottelay, Paul F., and T. Campbell-Copeland, *The Soldier in Our Civil War: Columbian Memorial Edition*, Stanley-Bradley Publishing Co., 1893, UNCLASSIFIED, in two volumes (475 pp and 444 pp, respectively) Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The data base contains 120 items covering actions from skirmishes to battles that occurred during the time of the American Civil War.

b. **Information Provided.** The actions are arranged in chronological order, with dates ranging from 1 June 1861 through 2 April 1865. Information on each action includes the page of the source from which the information is taken, its starting date, type of action (*e.g.*, skirmish, engagement, battle, *etc.*), its location (city and state), duration in hours, Federal and Confederate personnel strengths, killed, wounded, prisoners, and missing. An assessment of which side won is also provided. The casualty categories used are killed, wounded, taken prisoner, and missing.

c. **Situational Descriptors Included.** A brief descriptive phrase is usually provided in the original source, such as "Expedition from St. Louis up the Missouri River," "Fort Fillmore, New Mexico, abandoned," and the like. These were not carried forward into the electronic data base.

d. **Data Sources Used.** Various, but the chronological list of events is said to have been compiled by the coauthor (T. Campbell-Copeland) "from Official Records of the Federal and Confederate Armies, the Rebellion Record, Phisterer's Statistical Record, Hammersley's Army Register, the Official Army Register, Statistical volumes and pamphlets of the Department of the Interior, The Annual Cyclopædia, Twenty Years of Congress, and such other thoroughly reliable books and documents as have been made available."

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the actions and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Only those actions for which the source document provides estimates of the killed and wounded, as well as personnel strengths, for both sides are carried forward to the electronic data base. It is not always clear whether the dashes in the original source document actually represent zero quantities, and when they instead represent a category of "data not available." Accordingly, the dashes in this data base correspond to those in the source.

4. **Remarks.** To the extent that the data are trustworthy, this provides a unique compilation of data on many of the smaller actions that took place in the Civil War.

ORALFORE-1972

1. **General Description.** Dupuy, Trevor N., and Haynes, Grace P., "Opposed Rates of Advance of Large Forces in Europe (ORALFORE)," Historical Evaluation and Research Organization (HERO), 28 August 1972, UNCLASSIFIED, prepared by HERO for the US Army Deputy Chief of Staff for Military Operations (DAMO-PLW) under Contract No. DAHC19-72-C-0015, available from DTIC as AD-902 830. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The ORALFORE work was done (p 1, Introduction) "to provide insight into some of the basic problems of dealing with rates of advance in wargaming by attempting to ascertain whether historical combat experience provides a basis for calculating rates of advance of large forces (divisions and corps); by examining the assumption, accepted as valid in current models and wargames, that rates of advance can be related directly to the force ratios of opposing military forces; and by analyzing the factors that have influenced rates of advance in historical combat ..."

Specifically, the ORALFORE data were taken from those tables in the source entitled "Analysis of Rate of Advance of ...," where the ellipsis stands for the name of the unit and the period concerned, e.g., "7th Panzer Division, 10-24 May 1940." These tables give data on the 14 advances listed below.

List of operations considered in ORALFORE

Case no.	Short name	Description
1	7th Panzer	Ardennes-Flanders Offensive, Ger 7th Panzer Division (Rommel), 20-24 May 40
2	XLVII Corps	Orel to Moscow, Ger XLVII Corps, 13-30 Nov 41
3	18th Panzer	Orel to Moscow, Ger 18th Panzer Division, 13-23 Nov 41
4	29th Mtd	Orel to Moscow, Ger 29th Motorized Division, 23 Nov-3 Dec 41
5	XL Corps	Don River to Caucasus, Ger XL Panzer Corps, 21 Jul-23 Aug 42
6	3d Panzer	Don River to Caucasus, Ger 3d Panzer Division, 21 Jul-23 Aug 41
7	VII Corps	Argentan to Liege, US VII Corps, 13 Aug-12 Sep 44
8	3d Armored	Argentan to Liege, US 3d Armored Division, 13 Aug-12 Sep 44
9	1st Inf	Argentan to Liege, US 1st Infantry Division, 13 Aug-8 Sep 44
10	9th Inf	Argentan to Liege, US 9th Infantry Division, 15 Aug-10 Sep 44
11	XX Corps	Le Mans to Metz, US XX Corps, 14 Aug-13 Sep 44
12	7th Armored	Le Mans to Metz, US 7th Armored Division, 14 Aug-13 Sep 44
13	XII Corps	Saar (Lorraine) Campaign, US XII Corps, 8 Nov-7 Dec 44
14	4th Armored	Saar (Lorraine) Campaign, US 4th Armored Division, 9 Nov-7 Dec 44

The source tables contain somewhat more detail than we chose to transcribe into our data base. We omitted those parts presenting force ratios computed in various ways (these depend upon many

assumptions, extrapolations, and complicated procedures extraneous to the unvarnished historical record); estimates of how much of each day's advance distance was against negligible, slight, moderate, or intense opposition (the intensity categories are themselves subjective, and rating proportions of a day just compounds the subjectivity); air support present or absent (no evaluation of how much it was); logistics (data too sparse); and situation on flanks (given for advancing forces but not for defending forces). The source gives additional information in its day-by-day narrative accounts of the action, and in its other tables (such as orders of battle, lists of weather conditions, *etc.*). Unfortunately, it was impractical to compress all of this information into data base form.

b. Information Provided. The following information is tabulated for the operations cited in the previous paragraph. Operation name, route/locale/sector, date (values are given day by day), distance advanced, weather for armor, weather for air, terrain, road net, obstacles, opposition intensity, attacker and defender condition, attacker and defender mission, attacker and defender casualty fraction (in percent).

Here the following definitions apply. Weather for armor is a judgment of the suitability of weather for armor operations. Weather for air is a judgment of the suitability of weather for air operations. Terrain is a judgment of terrain mobility. Road net is a judgment of road net suitability for mobility. Obstacles are obstacles to advance, coded as follows: DM = Exceptionally effective demolitions, DS = Desert, FL = Flooded area, FZ = Fortified zone, R = River, SA = Sabotage by local populace, UR = Urban area. Opposition intensity is a judgment of the intensity of opposition to advance, coded as follows: I = Intense, M = Moderate, S = Slight, N = Negligible. Attacker and defender condition is a judgment of their readiness for operations, coded as follows: FR = Fresh, T = Tired, W = Weary, E = Exhausted. Attacker mission is the attacker's posture/mission, coded as follows (mixed entries in ORALFORE are coded as fractions. *I.e.*, ORALFORE entry 2-3 coded as 2.5): 1 = Unlimited, 2 = Related to specific objectives or contingencies, 3 = Related to adjacent unit's advance, 4 = Limited by higher authority control, 5 = Other. Defender mission is the defender's posture/mission, coded as follows: WD = Withdrawal, HD = Hasty defense, FD = Fortified defense, PD = Position defense, DL = Delay. Attacker and defender casualties are their losses in percent.

Casualties are identified as percent losses.

c. Situational Descriptors Included. These are as explained in the preceding paragraph.

d. Data Sources Used. US and German documents in US archives, and selected books. In some instances, gaps in the data were filled in by interpolation and/or extrapolation on the basis of professional military or historical experience.

e. Diskette Format. Diskette contains one file in spreadsheet form with one row for each of the 180 major units actions and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Daily percentage casualties to the attacker are provided on a nearly daily basis for all of the cases listed in paragraph 2a above. However, except for the last four cases (i.e., cases 11 through 14), the percentage casualties to the defending side are listed as "unknown." Hence, if the percentage casualties to both sides are required, only these last four cases provide the needed information.

4. **Remarks.** The defender's percentage casualties for cases 11 (XX Corps) and 12 (7th Armored) given in the source document are identical for 14 Aug 44 through 2 Sep 44. Presumably, this is a typographical or other error.

PARMISC-1993

1. **General Description.** Miscellaneous personnel attrition data assembled from various sources, 1993 UNCLASSIFIED. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The diskette contains one file in spreadsheet form detailing personnel strengths and losses in 209 battles.

b. **Information Provided.** The starting and ending dates, duration in fractions of a day (when known accurately enough to support such a determination), identification of the attacking and defending sides, personnel strengths and losses to each side, and a determination of which side was victorious.

c. **Situational Descriptors Included.** Not tabulated.

d. **Data Sources Used.** The data were assembled from the following sources:

(1) *America's First Battles: 1776-1965*, Charles E. Heller & William A. Stofft (eds), University Press of Kansas, 1986.

(2) *The Age of Voltaire*, Will & Ariel Durant, Simon and Schuster, NY, 1965.

(3) *Dictionary of Battles*, David Eggenberger, Thomas Y. Crowell Co., NY, 1967.

(4) *A Dictionary of Battles (1816-1976)*, Brigadier Peter Young, Mayflower Books, NY, NY 10022, 1978.

(5) *Kursk: The Clash of Armor*, Geoffrey Jukes, Ballantine's Illustrated History of WWII, Battle Book No. 7, Ballantine Books, NY, NY, 1968.

(6) *Strategy and Tactics*, No 49, Mar/Apr 1975.

(7) *History of US Naval Operations in World War II*, Samuel Eliot Morison, Little, Brown and Co., Boston, 1947-1960.

(8) *Victory in Papua*, Samuel Wilks, Office, Chief of Military History, Department of the Army, (Green Book series), 1957.

e. **Diskette Format.** The diskette contains one file in spreadsheet form, with rows for the battles. There are 261 rows. The columns provide space for the following items of information: sequence number (an arbitrary number used to identify the row); reference number (the number of the source from which the data was taken); page number (the number of the page of the source at which the information used begins); from which the information is taken; the name of the battle; the battle's

starting and ending dates; identification of the attacking and defending sides; the initial personnel strength and casualties to each side; a determination of which side won; and explanatory notes.

3. **Other and Miscellaneous.** Not used.

4. **Remarks.** Since some sources provided only a few cases, this miscellaneous data base was created to provide a vehicle for reporting those cases.

POGOGORO-1993

1. **General Description.** Adapted from Historical Evaluation and Research Organization report "Historical Scenarios of Soviet Breakthrough Efforts in World War II," April 1978, sponsored by the USAF Tactical Fighter Weapons Center, Nellis AFB, Nevada, under Contract No. F49620-77-C-0026. Data base on diskette furnished to DTIC concurrently with this document.
2. **Data Base Description.**
 - a. **General.** The data base provides the estimated Soviet and German personnel strengths and losses for each day of the Pogoroloye-Gorodische operation (4 August 1942 to 12 August 1942).
 - b. **Information Provided.** Information on each day includes the strength of each side at the start of that day, its losses during the day, and its strength at the end of the day. The casualty category used is losses.
 - c. **Situational Descriptors Included.** None, other than the day of the operation.
 - d. **Data Sources Used.** Various.
 - e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the days and columns for each of the items of data provided.
3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the report cited in paragraph 1 above.
4. **Remarks.** This provides some data on strengths and losses to both sides as a function of time into the operation, which are seldom reported. For other data bases purporting to provide such information, see CRETE-1993, INCHON-1993, IWOJIMA-1993, and WESTWALL-1993.

REISTERK-1986

1. **General Description.** Selected extracts from Reister, Frank A., *Battle Casualties and Medical Statistics: U.S. Experience in the Korean War*, The Surgeon General, Department of the Army, Washington, DC, 1986, 172 pp, UNCLASSIFIED, book available from libraries. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The data base consists of two separate files. The first contains information on division-level operations, and the second gives information on regimental-level operations.

b. **Information Provided.**

(1) The division-level file contains information on US Army division strengths and losses during 43 major operations of the Korean War. The major operations are classified by type as Offensive, Pursuit, Static, Limited, Defensive, and Withdrawal. These are defined as follows.

Offensive: a major offensive operation involving more than one division over a wide front and representing a major United Nations effort.

Pursuit: exploitation of complete disintegration of the North Korean Army in retreat.

Static: a type of operation developed from arbitrary phase lines which usually followed the natural features of terrain, such as river or ridge lines, and delineated major objectives on offense and composed secondary lines of resistance on defense.

Limited: limited operations with relatively small forces of regimental size or smaller, undertaken to keep the enemy off balance and to seize dominating terrain in order to improve the defensive position.

Defensive (Enemy Offensive): major offensive operations by North Korean and Chinese Communist forces.

Withdrawal: major withdrawal actions, both of which resulted from massive Chinese Communist intervention.

There is one line for each major operation. The columns give the page of the source from which the data are taken, the type of operation, the name of the operation and its starting and ending dates, the number of divisions taking part in the operation and the total number of division-days expended on it, the mean division strength, and the number of KIA, WIA-carded for record only, and WIA-admitted to a medical treatment facility.

(2) The regimental-level file contains information on US Army regimental losses in the major operations. It has 52 lines, each line giving the type of major operation involved, as defined above, the type of tactical action in which the regiment was involved (e.g., attack against strong

defense, defense against enemy counterattack, patrolling, *etc.*), the number of regiments involved, the number of regiment-days of effort devoted to the tactical action, mean strength of the regiments involved, and the number of KIA, WIA-CRO, and WIA-admitted.

(3) The casualty categories used are KIA, wounded in action and carded for record only, and wounded in action and admitted to a medical treatment facility.

c. **Situational Descriptors Included.** As specified by the type of operation and the type of tactical activity undertaken by a regiment.

d. **Data Sources Used.** Official US records of the Korean War.

e. **Diskette Format.** Diskette contains two files in spreadsheet form.

3. **Other and Miscellaneous.** The source from which the diskette files was adapted contains considerably more information of various kinds regarding losses to US Army forces during the Korean War. It was not practical during the scope of this project to reduce much of that information to electronic form.

4. **Remarks.** This is one of the handiest sources of information on the effect of tactical activity on losses.

RMC-1993

1. **General Description.** Cockrell, James K., and Carter, Donn, "Research Study on Predictive War Game Factors," RMC Research Corporation, unnumbered, March 1974, 251 pp, UNCLASSIFIED, available from the Technical Centre of the Supreme Headquarters, Allied Powers in Europe (SHAPE), prepared under SHAPE Technical Centre Contract C.72-03. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Includes 150 attack days by forces of brigade to division size against defenses in place (as opposed to delaying actions) during the first 18 days (16 December 1944 to 2 January 1945) of the Ardennes campaign of World War II (Battle of the Bulge), and a second set of 100 battle days from the Lorraine Campaign that occurred between 8 September and 27 November 1944 in eastern France.

b. **Information Provided.** The tabulated data include the sector of the front, the date, the number of infantry platoon equivalents, tank strength, antitank firepower, artillery firepower, personnel casualties, tank and tank-equivalent losses, visibility, mobility, attacker and defender posture, and the distance advanced. The following special definitions apply.

(1) **ATTACK TYPE** code is as follows: 0 = a US main attack as part of a continuing offensive; 1 = a German attack as part of a continuing offensive; 6 = a German attack combined with a US counterattack; 7 = a US attack combined with a German counterattack. **ATTACK DESCRIP** is an alphabetical description of the **ATTACK TYPE** code.

(2) **POSTURE** code is as follows: 5 = front line troops in well dug-in positions with at least some overhead cover—an organized, well-prepared defensive position; 6 = front line troops in foxholes with all heavy weapons dug in; 7 = some elements of the defense as in level 6, and a significant portion not well dug in or organized; 8 = a hasty defense in which there has not been opportunity to organize the defense or to dig in. This level was also used for "meeting engagements" (i.e., an attack meeting a counterattack, two attacks colliding, etc.).

(3) **ARTY US** is the number of rounds of US artillery ammunition expended in close (direct) support of that sector on that date. **ARTY GER** is coded as follows: 1 = low intensity of German artillery; 2 = moderate intensity of German artillery; 3 = high intensity of German artillery; 4 = very high intensity of German artillery.

(4) Four environment codes are provided. The first pair ("ENV LINE" or E1 and E2) are mobility restrictions and visibility ranges in the area of the defensive position. The second pair ("ENV PEN" or E3 and E4) are these same conditions along the axis of penetration. The codes used for E1

and E3 are: 0 = no restriction on infantry or tank movement; 1 = infantry unrestricted, tanks canalized; 2 = infantry unrestricted, tanks severely canalized; 3 = infantry unrestricted, tanks must breach; 4 = infantry canalized, tanks canalized; 5 = infantry canalized, tanks severely canalized; 6 = infantry canalized, tanks must breach; 7 = infantry canalized, tanks difficult to breach; 8 = infantry must breach, tanks must breach; 9 = infantry and tanks difficult to breach. The codes used for E2 and E4 are: 1 = not over 100 meters visibility range generally available; 2 = not over 300 meters visibility range generally available; 3 = not over 1,000 meters visibility range generally available; 4 = well over 1,000 meters visibility range generally available.

(5) TANK ATTN is the total armor loss rate per thousand armor vehicles per day. PERS ATTN is the personnel battle casualty rate per thousand personnel per day, computed on the basis of 40 men per infantry platoon.

(6) ADV DIST is measured in units of and to 0.1 km from the line occupied by the defense when the day's fighting began. That is, only penetration distance was counted, not advance-to-the-assault-position distance.

(7) FRONTAGE is an estimate of the length in kilometers of the part of the front line which was attacked in each battle day—not merely the points of close contact, but the length of the front the defending force had to protect as a direct result of the attacking force's actions and immediate capabilities.

(8) The casualty categories used are losses per thousand per day.

c. **Situational Descriptors Included.** Attack type code and description, posture code, code to identify which side was attacking or defending, US and Germany artillery, environmental conditions, number of infantry platoons on the US and German sides, quantity of US and German armor and antiarmor weapons, proportion of the combat task force that was mechanized, tank and personnel attrition measures, and width of front.

d. **Data Sources Used.** From p 7: "The US National Archives Captured Records Center was the primary source for German unit records and other documents. These included microfilmed copies of unit operational records, war maps, and results of interviews or responses from captured German division, corps, and army commanders and staff officers. ... For US battle day information the primary source was the National Records Center of the Archives. After action reports, daily and other periodic reports, journal files, histories, and order of battle studies from Army level down to battalion level were examined to establish the composition, location, disposition, mission and activity of the units of interest."

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 250 actions and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following studies:

a. Goad, Rex, "Prediction of Opposed Movement and Casualty Rates," Operations Research Division, SHAPE Technical Centre, The Hague, Professional Paper STC-PP-75, July 1973, 26 pp. UNCLASSIFIED, available from publisher and libraries.

b. Goad, Rex, "Predictive Equations for Opposed Movement and Casualty Rates for Land Forces," Operations Research Division, SHAPE Technical Centre, The Hague, Professional Paper STC-PP-90, August 1974, 47 pp, UNCLASSIFIED, available from DTIC (AD-921 726).

c. Schaefer, Alfred, "Zur Ermittlung Funktionaler Zusammenhänge Zwischen Bewegungsgeschwindigkeit und Verlustraten Militärischer Verbände im Gefecht [On the Determination of the Functional Relationship Between Speed of Movement and Loss Rates of Military Units in Combat]," Hochschule der Bundeswehr München, HSBwM-ID03/77, 9 June 1977, 240 pp, UNCLASSIFIED, available from publisher (Hochschule der Bundeswehr München).

4. **Remarks.** At the request of the SHAPE Technical Centre, the well-known British statistician, Dennis V. Lindley, then of the University College of London, reviewed the RMC report. His findings were reported to the SHAPE Technical Centre as an informal report addressed to Dr. Rex Goad and dated 24 January 1976, and copies can be obtained from him. Prof. Lindley abstained from a detailed review of the attrition data because "... a substantial amount of the attrition data is missing: furthermore, they appear to be missing in a systematic way, in the sense that the occasions when internal evidence suggests large losses are just those occasions where no casualty figures are available." He also expressed concern that RMC's statistical analysis of rates of advance suffered from overfitting (i.e., introducing too many terms into the regression equation) and pointed out that this will inevitably produce an R^2 value that is misleadingly high and not actually justified by the basic data. Clearly, RMC's attrition analysis suffers similar faults, so that further justification or confirmation from other sources would be highly desirable.

ROADATA-1990

1. **General Description.** Helmbold, Robert L., "A Compilation of Data on Rates of Advance in Land Combat Operations," U.S. Army Concepts Analysis Agency Research Paper CAA-RP-90-4, February 1990, UNCLASSIFIED. Report available from DTIC as AD-A220 426. Diskette available from DTIC as AD-M000 115 (two diskettes, UNCLASSIFIED, February 1990).

2. **Data Base Description.**

a. **General.** This data base includes over 7,000 entries tabulating reported rates of advance of land combat forces from ancient times to the present. The following descriptors are included when available: operation or battle name; route, locale, or sector; distance advanced; time duration of the advance; starting and ending dates and times; operational intensity; governing mode of movement (such as foot, horse, or motor); vegetation; footing; and force size.

b. **Information Provided.** To the extent it is available, the information on each instance of rate of advance includes name of the operation involved; the route, locale, sector, or other geographical location information; distance traveled; elapsed time; speed of travel; starting and ending dates; intensity of opposition; primary mode of movement; type of terrain topography traversed; vegetation; footing; force size; and comments. However, there are sizable gaps in the available information. The data seldom provide any information on losses.

c. **Situational Descriptors Included.** As given in paragraph 2h above.

d. **Data Sources Used.** Various, but in all cases were secondary sources such as reports, dictionaries of battles, books on military history, etc.

e. **Diskette Format.** The data base is provided on a set of two 5.25-inch 1.2 megabyte (high density) diskettes usable on IBM-compatible personal computers operating under MS-DOS 3.21 or higher. The diskettes contain the 36 files listed below. Note the slight departures from the list of primary data bases given in the hard copy documentation. First, in order to make each of the files small enough to fit in most personal computers, the Ardennes Combat Simulation Data Base data has been subdivided into seven files according to date. Second, the CAA data base of Battles—Version 1990 (CDB90) file has been renamed the CDB90RAT.WKS file because the full CDB90 data file was not used; only the movement rate data derived from it is given in the CDB90RAT.WKS file. Finally, even though it was not described in the hard copy documentation, one additional file (GIBBON.WKS) has been included; it is described in CAA-RP-3, cited in paragraph 3a below. Unless otherwise stated, all the following are in the form of spreadsheets readable by LOTUS 1-2-3TM and many other spreadsheet and data base programs.

- (1) README.TXT, an ASCII file describing the data base.
- (2) DEC15-19.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 15 Dec-19 Dec 44.
- (3) DEC20-24.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 20 Dec-24 Dec 44.
- (4) DEC24-29.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 24 Dec-29 Dec 44.
- (5) DEC30-J3.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 30 Dec 44-03 Jan 45.
- (6) JAN04-08.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 4-8 Jan 45.
- (7) JAN09-13.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 9-13 Jan 45.
- (8) JAN14-16.WKS, rates of advance from the Ardennes Combat Simulation Data Base for the period 14-16 Jan 45.
- (9) ALEXANDE.WKS, rates of advance for the campaigns of Alexander the Great.
- (10) ANDREWS.WKS, rates of advance in the early World War II North African campaigns.
- (11) BAORG.WKS, rates of advance of British infantry units in northwest Europe during World War II.
- (12) CAESAR.WKS, rates of advance for some of Caesar's campaigns.
- (13) CDB90RAT.WKS, rates of advance from the CDB90 data base of battles.
- (14) DESANTIS.WKS, rates of advance for selected World War II engagements.
- (15) EASTFRON.WKS, rates of advance for selected Eastern Front battles of World War II.
- (16) ENGELS.WKS, rates of advance for selected campaigns of Alexander the Great.
- (17) GIBBON.WKS, rates of advance for a cavalry unit during the American Indian wars of the post-Civil War period.
- (18) GLANTZ.WKS, rates of advance for selected World War II Eastern Front campaigns.
- (19) GUSTAVUS.WKS, rates of advance for selected campaigns from Gustavus Adolphus's era.
- (20) HANNIBAL.WKS, rates of advance for selected campaigns of the ancient Romans.

(21) HULSE.WKS, rates of advance for selected US Armored divisions in the northwest European theater during World War II.

(22) LONGMARX.WKS, rates of advance during the Communist Chinese "Long March" lead by Mao Tse-Tung.

(23) MISROAD.WKS, miscellaneous instances of rates of advance in selected military and civilian operations gathered from a variety of sources.

(24) NAPOLEON.WKS, rates of advance in selected Napoleonic campaigns.

(25) NORMANDY.WKS, rates of advance by US Army infantry division attacks in Normandy during World War II.

(26) ORALFORE.WKS, rates of advance by selected large forces on the Western and Eastern Fronts during World War II.

(27) PARSONS.WKS, rates of advance by US infantry battalions in selected World War II operations.

(28) QUICKWIN.WKS, rates of advance for selected campaigns in which one side gained a "quick win".

(29) RADZIEV.WKS, rates of advance by Soviet tank armies in selected World War II operations.

(30) RECORD.WKS, unusually high rates of advance achieved by armored forces in World War II or in the Arab-Israeli wars.

(31) RMC.WKS, rates of advance by selected German or US divisions during the World War II Ardennes and Lorraine campaigns.

(32) ROWLAND.WKS, rates of advance in selected unopposed movements from the US Civil War, Boer War, World War I, World War II, and Falklands War.

(33) SAVKIN.WKS, rates of advance for the major Soviet Red Army operations of World War II.

(34) SIEGFRIE.WKS, rates of advance by US infantry divisions during the World War II Siegfried Line campaign.

(35) WAINSTEI.WKS, rates of advance in selected operations from the Franco-Prussian War through the Arab-Israeli wars (but the vast majority of the cases are from World War II).

(36) XENOPHON.WKS, rates of advance on the "March of the Ten Thousand" Greek mercenary troops during their escape from the Persian Empire circa 400 BC.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following papers. They, together with the hard copy documentation, are indispensable to a proper understanding of the origins, uses, limitations, and proper interpretation of the data contained in these data base files.

a. "A Survey of Past Work on Rates of Advance in Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-3, February 1990 (UNCLASSIFIED, available from DTIC as AD-A217 891).

b. "Rates of Advance in Historical Land Combat Operations," US Army Concepts Analysis Agency Research Paper CAA-RP-90-1, June 1990 (UNCLASSIFIED, available from DTIC as AD-A225 635).

4. **Remarks.** This data base of rates of advance is unique in terms of its completeness. It makes available several sets of rate of advance data that previously were obtainable only from obscure and hard-to-get sources. However, much of the data on such things as topography, footing, intensity of opposition, *etc.*, are not available and so could not be provided.

SP128-1961

1. **General Description.** Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group, Staff Paper CORG-SP-128, 1 July 1961, 178 pp, UNCLASSIFIED. Report available from DTIC as AD-480 975. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Includes 92 battles from 1741 to 1945.

b. **Information Provided.** Information on each battle includes the battle name; year in which the battle took place; a key to the source of the information provided; identification of the attacker and defender nationalities; initial personnel strengths of the attacker and defender; casualties to the attacker and defender; how the source identified the type of casualties reported (*i.e.*, wounded, killed, "losses," and so forth); duration of the engagement in days (if known); and identification of which side won. The casualty categories used are L (losses), M (killed, wounded and missing), W (killed and wounded), and NA (not identified as to type).

c. **Situational Descriptors Included.** None, other than the name and year of the engagement.

d. **Data Sources Used.** Various, but in each case, secondary sources such as encyclopedias, dictionaries of battles, books on military history, *etc.*

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 92 battles and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following papers.

a. "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group Staff Paper CORG-SP-128, 1 July 1961. AD-480 975.

b. "Lanchester's Equations, Historical Battles, and War Games," Proceedings of the Eighth Military Operations Research Symposium, Pt. Mugu, CA, 1961.

c. "Historical Data and Lanchester's Theory of Combat: Part II," CORG-SP-190, August 1964. AD-480 109.

d. "Some Observations on the Use of Lanchester's Theory for Prediction," Opns. Res., 12, pp 778-781, September-October 1964.

e. "Problems in Testing Combat Models," presented to the First National Meeting, Military Applications Section, ORSA, Houston, TX, 1965.

- f. "A Modification of Lanchester's Equations," *Opns. Res.*, 13, pp 857-859, September 1965.
 - g. "Some Observations on Validating Combat Models," Presented to the NATO Conference on Recent Developments in Lanchester Theory, Munich, Germany, 3-7 July 1967.
 - h. "Probability of Victory in Land Combat as Related to Force Ratio," RAND P-4199, October 1969. AD-696 489.
 - i. "Air Battles and Ground Battles—A Common Pattern?," RAND P-4548, January 1971. AD-718 975.
 - j. "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," RAND R-772-PR, June 1971. AD-729 769.
 - k. "Some Observations on the Choice of Exponent in Lanchester's Law," presented to the Ad Hoc Working Group on the Theory of Combat, cosponsored by DNA and NPGS, June 1979.
4. **Remarks.** This data base was created to support one of the earliest attempts to apply the parameters suggested by Lanchester's square law of combat attrition to an examination of historical battle data. Some misprints in the source have been corrected in the data base version.

SP190-1964

1. **General Description.** Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat: Part II," Combat Operations Research Group Staff Paper CORG-SP-190, August 1964, 132 pp, UNCLASSIFIED. Report available from DTIC as AD-480 109. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** Includes 83 battles from 280 BC to 1944 AD.

b. **Information Provided.** Information on each battle includes the battle name; the war, campaign, or article of which this battle is a part; a key to the source of information provided; the year in which the battle took place; identification of the principal nationality on the attacker and defender sides; initial personnel strengths of the attacker and defender; casualties to the attacker and defender; how the source identified the type of casualties reported (i.e., killed, wounded, and so forth); duration of the engagement in days (if known); and identification of which side won. The casualty categories used are L (losses), M (killed, wounded and missing), W (killed and wounded), and NA (not identified as to type).

c. **Situational Descriptors Included.** None, other than the name and year of the engagement.

d. **Data Sources Used.** Various, but in all cases were secondary sources such as encyclopedias, dictionaries of battles, books on military history, etc.

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 83 battles and columns for each of the items of data provided.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following papers.

a. "Historical Data and Lanchester's Theory of Combat: Part II," CORG-SP-190, August 1964. AD-480 109.

b. "Some Observations on the Use of Lanchester's Theory for Prediction," Opns. Res., 12, pp 778-781, September-October 1964.

c. "Problems in Testing Combat Models," Presented to the First National Meeting, Military Applications Section, ORSA, Houston, TX, 1965.

d. "A Modification of Lanchester's Equations," Opns. Res., 13, pp 857-859, September 1965.

e. "Some Observations on Validating Combat Models," presented to the NATO Conference on Recent Developments in Lanchester Theory, Munich, Germany, 3-7 July 1967.

f. "Probability of Victory in Land Combat as Related to Force Ratio," RAND P-4199, October 1969. AD-696 489.

g. "Air Battles and Ground Battles—A Common Pattern?", RAND P-4548, January 1971. AD-718 975.

h. "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," RAND R-772-PR, June 1971. AD-729 769.

i. "Some Observations on the Choice of Exponent in Lanchester's Law," presented to the Ad Hoc Working Group on the Theory of Combat, cosponsored by DNA and NPGS, June 1979.

4 **Remarks.** This data base provided a separate sample of data that was used to confirm the findings obtained from the earlier study of the SP128 data base. Some misprints in the source are corrected in the data base version.

VOEVODSK-1993

1. **General Description** Voevodsky, John, "Quantitative Analysis of Nations at War," Peace Research Reviews, vol 4, 1970, pp 1-63, UNCLASSIFIED, published by Canadian Peace Research Institute, Oakville, Ontario, available from libraries. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** This data base contains strength and loss data at a function of time into each of the following wars: US Civil War, World War I, World War II, the Korean War, and the Vietnam War.

b. **Information Provided.** Information on each war includes the name of the war, the month and year to which the data apply, the strength of the in-theater force, and the cumulative losses to date for battle casualties and battle deaths. The casualty category used is identified as encompassing battle casualties and battle deaths. Battle casualties can generally be interpreted as including the KIA, WIA, and CMIA. Battle deaths can generally be interpreted as including the KIA and DOW.

c. **Situational Descriptors Included.** None, other than the name and year of the war.

d. **Data Sources Used.** Various, but in most cases were sources deemed reliable, such as US Department of Defense information and standard references.

e. **Diskette Format.** The data base consists of one file in spreadsheet form, with separate sections devoted to each war. Within each war, the date, strength, and losses are tabulated.

3. **Other and Miscellaneous.** Examples of the use of this data base may be found in the following studies:

a. Voevodsky, John, "Quantitative Behavior of Warring Nations," Systems Analysis Division, Office of the Chief of Naval Operations, Department of the Navy, 17 March 1968, UNCLASSIFIED.

b. Voevodsky, John; and Waleszczak, S. L., "Quantitative Analysis of the Vietnam War Through 30 June 1969," submitted to the Military Operations Research Society for General Session Consideration, 27 February 1970, UNCLASSIFIED.

c. Voevodsky, John, "Crisis Waves: The Growth and Decline of War-Related Behavioral Events," The Journal of Psychology, Vol 80, 1972, pp 289-308, UNCLASSIFIED, available from libraries.

d. Voevodsky, John, "Quantitative Behavior of Warring Nations," The Journal of Psychology, Vol 72, 1969, pp 269-292, UNCLASSIFIED, available from libraries.

c. Voevodsky, Jonn, "A Mathematical Representation of Time-Variant Quantitative Behavior," The Journal of Psychology, Vol 68, 1968, pp 129-140, UNCLASSIFIED, available from libraries.

4. **Remarks.** This collects together in one place data that otherwise is scattered and often hard to locate. This facilitates comparing the data on one war with another.

WALKER-1991

1. **General Description.** Walker, Stephen K., "Armored Vehicle Weapon Impact Assessment in Southeast Asia," Air Force Institute of Technology Thesis, AFIT/GOR/ENS/91M-20, 290 pp, UNCLASSIFIED/LIMITED. Document available from DTIC as AD-B156 495. Diskette available from DTIC as "Walker Crew Casualty Data Base," AD-M200 173L.

2. **Data Base Description.**

a. **General.** The data base was generated from historical data on 706 vehicle damage incidents recorded in the document files associated with the GNVHSEADB data base maintained at the Survivability/Vulnerability Information Analysis Center (SURVIAC). These data were originally collected in the field by US Army teams as part of the Battle Damage Assessment Reporting Program (BDARP). Since each hit of a multiple hit was recorded separately, the final dBASETM data base contains 737 records, with records recording multiple hits cross-referenced via the Document Accession Number (DAN) to the other records on the same vehicle. These records address casualties to vehicle crewmen during the Vietnam War. The data cover the period from July 1969 to July 1970.

b. **Information Provided.** The casualty categories used are KIA, WIA, DOW, and MIA. Casualties are also classified as Not Wounded, Lightly Wounded (self-aid required), Moderately Wounded (unit aid man required), Severely Wounded (evacuation required), Dead, and Position Not Manned. [However, in the second classification scheme, the Lightly and Moderately Wounded categories were combined because the data did not permit distinguishing them.] The date and time of day of each incident are recorded whenever available.

c. **Situational Descriptors Included.** Each data base record consists of 72 fields. Several descriptors are provided when available, including such items as country where the incident occurred (South Vietnam or Cambodia), unit and branch of service to which the vehicle was assigned, vehicle speed, number of hits on the vehicle, azimuth and elevation of incoming rounds, effect of the hits on mission completion, and several others.

d. **Data Sources Used.** The GNVHSEADB data base maintained at the SURVIAC. These data were originally collected in the field by US Army teams as part of the BDARP.

e. **Diskette Format.** The diskette contains two files (SEASIA.DBF and SEASIA.DBT) for use with the dBASE program. The data base structure is as follows, where the entry headings give the dBASE field name (e.g., MODEL), the field type (e.g., N for "numeric" and C for "character"), the field width (in maximum number of spaces allowed), and explanatory notes. For example, the first field is named FILE_NUM. It is a numeric field, and has a width of 3 spaces.

- (1) FILE_NUM, N, 3.- Numeric designator of the file cabinets containing the paper records. Equal to 1 for all records in this data base.
- (2) DAN, N, 5.- Document Accession Number, a unique incident identification by which the original paper record file may be retrieved.
- (3) SERVICE, C, 4.- Branch of the military service that owned the vehicle (USA for all records in this data base).
- (4) MATRL_TYPE, C, 12.- The general type of vehicle involved in this incident.
- (5) MODEL, C, 3.- Coded vehicle designator.
- (6) MDL_DECODE, C, 15.- Vehicle model/series involved in this incident.
- (7) SER_NBR, C, 9.- Designated vehicle serial number.
- (8) INCID_CAT, C, 4.- The type of damage incident, which is CBT (combat) for all records in this data base.
- (9) ULTIM_DISP, C, 19.- Disposition of the vehicle after damage assessment.
- (10) DATE, D, 8.- Date in standard System 200 date format (month/day/year).
- (11) SORT_DATE, N, 6.- Date information in year/month/day (YY/MM/DD) format.
- (12) GMT, N, 4.- Greenwich Mean Time of the incident in four-digit, 24-hour clock system.
- (13) TIME, N, 4.- Local clock time of the incident in four-digit, 24-hour clock system.
- (14) PRI_UNIT, C, 10.- The Army division/regiment to which the vehicle was assigned for the operation.
- (15) SEC_UNIT, C, 10.- The Army battalion/company to which the vehicle was assigned for the operation.
- (16) MSN_GEN, C, 12.- The vehicle's general or first-level tactical mission description.
- (17) MSN_SPEC, C, 15.- The vehicle's second-level tactical mission description.
- (18) MSN_STAGE, C, 12.- The phase of the mission during which the incident occurred.
- (19) TGT_TYPE, C, 8.- Target type the vehicle was being used against.
- (20) CTRY, C, 4.- The country where the incident occurred.
- (21) VEH_SPEED, N, 5.- Vehicle speed at time of incident.
- (22) DIR_HITS, C, 4.- Number of rounds hitting the vehicle.
- (23) IND_HITS, C, 4.- Number of near misses (indirect hits) that damaged the vehicle.

- (24) OPS_IMPACT, C, 19.- Brief description of the vehicle action after being hit.
- (25) NBR_CAS, N, 2.- Total number of casualties resulting from the incident, whether crew members or not.
- (26) RPR_STATUS, C, 13.- Operational status of the equipment following a hit.
- (27) REPAIR_LVL, C, 10.- Organization or level primarily involved in repairing the vehicle.
- (28) REPAIR_MH, N, 5.- Estimated or total number of manhours to repair the vehicle.
- (29) DOWNTIME, N, 5.- Number of hours the vehicle is down for repairs.
- (30) FOTO, C, 3.- Information on whether repair or damage photos are included in the paper records of the incident.
- (31) NARRATIVE, C, 3.- Information on whether additional narrative material is included in the paper records for the incident.
- (32) EQMTSKETCH, C, 3.- Information on whether the vehicle diagram showing damaged areas and projectile direction is included in the paper records for this incident.
- (33) SCENARIO, C, 3.- Information on whether a diagram of the mission profile including the relation between the vehicle, its target, other friendly forces, and enemy positions is included in the paper records for this incident.
- (34) DAT_SOURCE, C, 5.- Indicates the level of indexing previously done on the data.
- (35) THREAT_CL, C, 15.- A first-level description of the threat weapon.
- (36) THRT_TYPE, C, 15.- A second-level description of the threat weapon.
- (37) ROUND_TYPE, C, 15.- Projectile type (armor piercing, high explosive, high explosive antitank).
- (38) AZIMUTH, C, 15.- The incoming direction of the projectile in azimuth.
- (39) ELEV, C, 15.- The incoming direction of the projectile in elevation.
- (40) HIT_LOC, C, 33.- One- or two-word description of where the hit was located.
- (41) THRT_EFFECT, C, 25.- The direct or indirect effect of the hit on the vehicle.
- (42) SYS_DMG, C, 40.- Ground vehicle systems, subsystems, and major components that were damaged.
- (43) SURV_MOD, C, 30.- Any survivability/vulnerability feature or modification that offset or negated the threat effects.

(44) SYS_PROT, C, 14.- The vehicle system, subsystem, or component protected by the survival modification described immediately above.

(45) PERS_FUNCT, C, 15.- The major title, duty, or responsibility of the individual involved as a casualty.

(46) PER_STATUS, C, 15.- The casualty status of the individual involved (DOW, KIA, MIA, or WIA).

(47) PRT_DEVICE, C, 18.- The protective device worn by the injured person.

(48) WOUND_LOC, C, 15.- A numeric code indicating the major anatomical region wounded (head and neck, thorax, abdomen, pelvis, legs, arms, total body).

(49) EVAC, C, 15.- Indicates whether or not the individual was evacuated from the incident location.

(50) XREF, C, 15.- Indicates the DAN for other incidents which occurred in the same engagement as the record being viewed. In the case of multiple hits, indicates the DAN number of the primary record for the vehicle hit.

(51) REMARKS, M, 10.- Memo field containing narrative information about the incident.]

(52) HULL_TURR, N, 1.- A code for the place on the vehicle impacted.

(53) HEX_AZMTH, N, 1.- A code for the general azimuthal direction of attack.

(54) HEX_ELEV, N, 1.- A code for the general elevation of the attack.

(55) DRV_PROT, N, 1.- Indicates whether the driver was wearing body armor at the time of the incident.

(56) DRVWNDLOC, N, 1.- A code for the major anatomical region of the driver's most severe wound.

(57) DRVWNDSEV, N, 1.- A code for the severity of the driver's wounds.

(58) TC_PROT, N, 1.- Indicates whether the track commander was wearing body armor at the time of the incident.

(59) TCWNDLOC, N, 1.- A code for the major anatomical region of the track commander's most severe wound.

(60) TCWNDSEV, N, 1.- A code for the severity of the track commander's wounds.

(61) GNR_PROT, N, 1.- Indicates whether the gunner was wearing body armor at the time of the incident.

(62) GNRWNDLOC, N, 1.- A code for the major anatomical region of the gunner's most severe wound.

(63) GNRWNDSEV, N, 1.- A code for the severity of the gunner's wounds.

(64) LDR_PROT, N, 1.- Indicates whether the loader was wearing body armor at the time of the incident.

(65) LDRWNDLOC, N, 1.- A code for the major anatomical region of the loader's most severe wound.

(66) LDRWNDSEV, N, 1.- A code for the severity of the loader's wounds.

(67) PERF, N, 1.- Indicates whether or not the hull was perforated.

(68) VEH_MOD, N, 1.- Indicates whether or not the vehicle was modified for survivability.

(69) MOB_KILL, N, 1.- Indicates whether or not the damage to the vehicle rendered it incapable of moving for at least 15 minutes, even if the crew could repair it in a longer period of time.

(70) FP_KILL, N, 1.- Indicates whether or not the damage to the vehicle rendered its main armament incapable of operation for at least 15 minutes, even if the crew could repair it in a longer period of time.

(71) CAT_KILL, N, 1.- Catastrophic kill. Indicates whether or not a massive and total destruction of the vehicle, usually accompanied by fire or explosion, occurred.

(72) CAS_IMPACT, N, 1.- Combined impact of casualties and vehicle damage on mission completion.

3. **Other and Miscellaneous.** The data base includes 104 incidents involving the M48 tank, 345 involving the M113 armored personnel carrier, 89 involving the M551 armored reconnaissance/airborne assault vehicle, 77 involving nine different types of trucks, and 91 involving 24 other assorted vehicles such as the M132 flamethrower, M577 command post, M43 duster, D7 bulldozer, arc-welders, and cement mixers. The data covers equipment assigned to the 25th, 1st, and 4th Infantry Divisions, the Americal Division, and the 11th Armored Cavalry Regiment.

4. **Remarks.** The author notes several considerations bearing on the interpretation of his findings. Among them are the following.

a. Classified sources were intentionally excluded in order to keep the study report UNCLASSIFIED.

b. Although variants of two of the vehicles studied in this work were still in service in the US Army at the time the source was prepared, none of these vehicles were exactly like those studied. Further, the materials currently used in US vehicles and the spectrum of threats they face can be quite different from those in the population studied.

c. Data on the use or presence of body armor was rarely collected by the BDARP teams, except for crewmen who were injured. This possible bias was considered in the author's analysis, and should be considered in any subsequent analysis.

d. We note that the CREWCAS-1991 data base provides some data on crew casualties in World War II.

WESTWALL-1993

1. **General Description.** Adapted from data in Historical Evaluation and Research Organization (HERO) Report Number 129, "Combat History Analysis Study Effort (CHASE) Data Enhancement Study (CDES)," 31 January 1986, in five volumes, UNCLASSIFIED, prepared for CAA under Contract Number MDA903-85-C-0252 and available from DTIC under the following accession numbers: AD-A175-712, AD-A175-713, AD-A175-714, AD-A175-715, and AD-A175-716; and in unnumbered HERO Report "Data Base Error Correction (DBEC)," 23 January 1987, UNCLASSIFIED, prepared for CAA under Purchase Order Number MDA903-86-M-8560 and available from DTIC under accession number AD-A176-750. Data base on diskette furnished to DTIC concurrently with this document.

2. **Data Base Description.**

a. **General.** The data base provides estimates of the strengths, reinforcements, and losses as a function of time during the battle between German and US forces during the World War II battle of the Westwall (2 October 1944 to 7 October 1944).

b. **Information Provided.** The battle is divided into 11 time periods of varying length. For each period, the strengths at its start, the losses and reinforcements during, and the strengths at the end are provided. The casualty category used is losses.

c. **Situational Descriptors Included.** None, other than the time period.

d. **Data Sources Used.** Various, but in most cases based on archival records.

e. **Diskette Format.** Diskette contains one file in spreadsheet form with one row for each of the 11 time periods and columns for each of the items of data provided.

3. **Other and Miscellaneous.** No known applications of this data base are available as of this writing.

4. **Remarks.** This provides some insight into strengths and losses during the progress of the action, which is not often reported. For other data bases purporting to provide such data, see CRETE-1993, INCHON-1993, IWOJIMA-1993, and POGOGORO-1993.

APPENDIX C

THE ICPSR DATA BASES

C-1. INTRODUCTION. This appendix describes some data bases of potential interest to military analysts that are available through the Inter-university Consortium for Political and Social Research (ICPSR), maintained by the Institute for Social Research, the University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106, Phone: 313-764-2570, FAX: 313-764-8041. ICPSR can be reached over the Internet at the address icpsr_netmail@um.cc.umich.edu. To reach the ICPSR Computer Support Group directly, the address is support@icpsr.umich.edu. The material in this appendix has been adapted from the ICPSR's *Guide to Resources and Services, 1993-1994*.

C-2. GENERAL DESCRIPTION OF THE ICPSR ARCHIVES

a. The Archive of ICPSR receives, processes, and distributes machine-readable data on social phenomena occurring in over 130 countries. The Archive contains machine-readable records of individual attitudes and social experience relevant to the full range of social science disciplines. The content of the Archive extends across economic, sociological, historical, organizational, social, psychological, and political concerns. Also included is an array of data on national attributes, domestic and international events, conflicts, international organizations, nation dyads, alliances, and international systems. The data resources of ICPSR are developed and maintained by the Archive, which performs two basic functions. First, the Archive acquires data and transforms them into forms that can be used by persons not intimately familiar with the original studies. Not all data collections received by the Archive are extensively "cleaned" in this manner. Rather, those major studies deemed by a set of advisory committees most likely to be of widespread interest are cleaned and fully checked for any problems that might inhibit their extended use. A classification system is used to inform the potential user of the extent to which particular data sets have been cleaned and documented by ICPSR staff. The Archive's second function is to disseminate data.

Currently (1993) the Archive has over 30,000 machine-readable files in nearly 2,900 study titles. This number continues to grow as nearly 200 titles are added to the holdings annually. All ICPSR machine-readable data resources are maintained and distributed in a manner designed to maximize utilization for secondary analysis. Membership in ICPSR is not a prerequisite for access to the data resources. However, virtually all of the data resources are available to individuals at ICPSR member institutions without charge, while the analogous services are available on a charge basis to individuals who are not so affiliated. All requests for data from individuals at member institutions should be made through the "Official Representative." Requests for data from individuals at nonmember institutions should be addressed to the ICPSR Director of Technical Services. Estimates of the costs involved in

providing specific data services will be provided upon request. No services will be provided to individuals at nonmember institutions until the cost has been agreed upon.

b. As of 1993, the ICPSR data base holdings are organized into the 17 major categories listed below. However, those in category III (Conflict, Aggression, Violence, Wars) are perhaps of greatest interest to most military operations analysts.

I. Census Enumerations. Historical and Contemporary Population Characteristics. (United States; Nations Other Than the United States).

II. Community and Urban Studies. (Studies of Local Politics; Detroit Area Studies; Historical Urban and Community Structure Studies).

III. Conflict, Aggression, Violence, Wars. (Conflict and Stability Within Nations; Conflict Between and Among Nations).

IV. Economic Behavior and Attitudes. (Continuing Series of Consumer Surveys; Surveys of Economic Attitudes and Behavior; Historical and Contemporary Economic Processes and Indicators).

V. Education. (United States; Nations Other Than the United States).

VI. Elites and Leadership. (United States; Nations Other Than the United States).

VII. Geography and Environment.

VIII. Governmental Structures, Policies, and Capabilities. (Behavior and Attitudes of Bureaucrats; Historical and Contemporary Public Policy Indicators and Outputs; Statistics on Government Operations).

IX. Health Care and Health Facilities.

X. Instructional Packages and Computer Programs.

XI. International Systems: Linkages, Relationships, and Events. (Dyadic and Small Group Interaction (Events Data); International Organizations; Structural Characteristics of the International System; Alliances and Military Affairs).

XII. Legal Systems.

XIII. Legislative and Deliberative Bodies. (Historical and Contemporary Roll Call Voting Records; Studies of Decisionmaking in Deliberative Bodies).

XIV. Mass Political Behavior and Attitudes. (Historical and Contemporary Electoral Processes; Political Participation; Public Opinion on Political Matters; Political Parties).

XV. Organizational Behavior. (United States; Nations Other Than the United States).

XVI. Social Indicators. (United States; Nations Other Than the United States).

XVII. Social Institutions and Behavior. (Minorities and Race Relations; Religion; Socialization, Students, and Youth; Age and the Life Cycle; Crime and the Criminal Justice System; Leisure and Recreation; Vital Statistics; Family and Gender).

c. Each ICPSR data base is described in the ICPSR guide according to the following format (although not all collections have entries for all of the fields described below). The following conventions are used to present technical information: (1) One-part data collections have one entry; (2), Small, multipart collections have individual entries; (3) Large, multipart collections show only a range of figures for each element, using one or (if there are several groupings of data) more entries.

(1) **Principal Investigator.** Each description begins with the name of the Principal Investigator (an individual or individuals or an agency), the authoritative title of the collection with qualifications specific to the current release of the data in parentheses if needed, and the ICPSR study number to be used when ordering.

(2) **Summary.** This is an abstract of the collection describing its purpose and substance.

(3) **Class.** Contains the ICPSR class number.

(4) **Universe.** Identifies the group of entities which comprise the object of study in a research investigation: the "population" of elements from which a sample is selected.

(5) **Sampling.** Describes the procedures used to extract the sample from the universe.

(6) **Note.** Used to describe details of interest to users.

(7) **Extent of Collection.** Lists the components of the entire collection, including the number of data files as well as files of machine-readable documentation and/or other material.

(8) **Data Format.** Lists all of the collection's available data formats.

(9) **Extent of Processing.** Describes processing steps performed on the collection, either by ICPSR, by the principal investigator, or by the data producer. The categories and abbreviations used for this are explained below.

All ICPSR data collections have been checked to determine that data and documentation correspond. ICPSR currently employs two systems to describe the extent of any additional processing a data collection has received. The first is a scheme of data "classes" that categorizes collections according to the level of processing performed by ICPSR. Class numbers appear at the end of the SUMMARY field in data collection descriptions. These class numbers are explained below.

Class I: Class I data sets have been checked, corrected if necessary, and formatted to ICPSR specifications. Also, the data may have been recoded and reorganized in consultation with the investigator to maximize their utilization and accessibility. A code book, often capable of being read by

a computer, is available. This code book fully documents the data and may include descriptive statistics such as frequencies or means. For those Class I data sets with code books available in printed form, a copy of the printed code book is supplied routinely to each "Official Representative." All Class I studies are available in multiple technical formats: SPSS control cards are also available for many Class I collections.

Class II: Class II studies have been checked and formatted to ICPSR specifications. Most non-numeric codes have been removed. The studies in this class are available in multiple formats, with SPSS control cards available for many Class II collections. The documentation exists as either a machine-readable code book (which may be edited and updated as required by further processing), a multilithed draft version, or a photocopy of the investigator's code book. Any peculiarities in the data are indicated in the documentation. A copy of the documentation is supplied when the data are requested.

Class III: Class III studies have been checked by the ICPSR staff for the appropriate number of data records and accurate data locations as specified by the investigator's code book. Often, frequency checks on these data have been made. Known data discrepancies and other problems, if any, will be communicated to the user at the time the data are requested. One copy of the code book for these data is supplied when the data are requested. The data themselves usually are available only in the format provided by the principal investigator.

Class IV: Class IV studies are distributed in the form received by ICPSR from the original investigator. The documentation for Class IV studies is reproduced from the material originally received. One copy of the documentation is supplied upon request from the Official Representative.

The second system, implemented in 1992 and intended ultimately to replace the system of data classes, describes discrete processing steps performed on the data, either by ICPSR, by the principal investigators of a collection, or by the data producers. This system presents the data processing information in the form of abbreviations in a separate field of the data collection description called EXTENT OF PROCESSING. The following is a key to these abbreviations:

BLANKS	= Data contain blanks.
CONCHK.ICPSR	= Consistency checks performed by ICPSR.
CONCHK.PI	= Consistency checks performed by Principal Investigator.
CONCHK.PR	= Consistency checks performed by Data Producer.
FREQ.ICPSR	= Frequencies provided by ICPSR.
FREQ.PI	= Frequencies provided by Principal Investigator.
FREQ.PR	= Frequencies provided by Data Producer.
MDATA	= Missing data codes standardised within the collection.
NONNUM	= Data contain nonnumeric codes.
RECODE	= ICPSR performed recodes and/or recalculated derived variables.
UNDOCCHK.ICPSR	= Checks for undocumented codes performed by ICPSR.
UNDOCCHK.PI	= Checks for undocumented codes performed by Principal Investigator.
UNDOCCHK.PR	= Checks for undocumented codes performed by Data Producer.

(10) Part Number. Gives the official number of the part and a descriptive name for that part for multipart collections.

(11) File Structure. Describes the structure of the part: rectangular, hierarchical, hierarchical in variable blocked format, or relational.

(12) Cases. Indicates the number of units of analysis.

(13) Records per Case. Gives the number of 80-character card image records per case when the data are in card image format. Gives the number of records per case when the data are available in logical record length format.

(14) Variables. Gives the number of variables contained in each file.

(15) Record Length. Gives the number of characters in each physical data record. Where alternate data formats are available, Record Length refers to the data in its logical record length or OSIRIS format. Record length for card image records is always 80 characters.

(16) Related Publication. Contains citations to important publications which are based on the data.

d. As a general practice, all data supplied by ICPSR will be routinely transmitted on magnetic tape supplied by the user. Users interested in data in forms suitable for use on microcomputers are encouraged to implement communications software that allows local "downloading" from mainframe equipment (or minicomputers) to microcomputers. Where implementing downloading capabilities may not be feasible locally, the ICPSR can supply direct data services for microcomputer applications. Users should contact Member Services for further information about procedures and charges for these services.

To order data, send either an ICPSR Data Request Form or a letter indicating the data sets desired to:

Member Services
ICPSR
Institute for Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
USA
Phone: 313-763-5010

Magnetic tape is needed for all data requests, unless the request is for data analysis only. Members may purchase a 2,400-foot tape from the Consortium for \$22; an intent to have ICPSR supply the tape should be stated at the time of the request.

e. Currently the ICPSR archive has over 30,000 machine-readable files in approximately 2,900 study titles (a rough estimate is that this represents over 6 million variables). This number continues to grow as nearly 200 titles are added to the holdings annually. Without some automated assistance, it

is impossible to gain a good understanding of what is in the holdings, especially at the variable level. Accordingly, ICPSR has developed four informational data bases in SPIRES (Stanford Public Information Retrieval System) to supply this need, all of which are available for general use. Three of these computer based resources describe ICPSR data holdings and will aid in identifying data needed for research, instructional, and other applications. The fourth contains bibliographic references to published and unpublished materials bearing upon survey research methodology.

f. Few of the ICPSR data sets are directly related to attrition rates in historical land combat operations. However, several are related to current matters of considerable interest to military operations research analysts involved in the analysis of the relation between military operations and political crisis situations.

C-3. SPECIFIC DATA BASES. The ICPSR guide lists the following items under its Category III.

a. ICPSR Data Bases Listed Under Category III (Conflict, Aggression, Violence, Wars), Subcategory A (Conflict and Stability Within Nations). Full details are given for only one item in this subcategory, otherwise only the title and ICPSR catalog number are given. See the ICPSR catalog for full descriptions of all items.

- (1) Domestic Conflict Behavior, 1918-1966 (ICPSR 5003).
- (2) Justifying Violence: Attitudes of American Men 1969 (ICPSR 3504).
- (3) Data Bank of Assassinations, 1948-1967 (ICPSR 5208).
- (4) Data Bank of Minority Group Conflict, 1955-1965 (ICPSR 5209).
- (5) Political Events Project, 1948-1965 (ICPSR 5206).
- (6) Systemic Conditions of Political Aggression (SCOPA) Project, 1955-1964 (ICPSR 5207).
- (7) Causal Model of Civil Strife, 1961-1965 (ICPSR 5009).
- (8) Civil Strife Events, 1955-1970 (ICPSR 7531).
- (9) Civil Strife Conflict Magnitudes, 1955-1970 (ICPSR 7485).

Principal Investigator. Gurr, Ted Robert, and Vaughan Bishop.

Summary. This collection consists of two files of data containing variables measuring the magnitude and intensity of political civil conflict during the period from 1955 to 1970 for 115 countries. Not all countries have data for all years. Measures were derived by aggregating and weighting coded information on specific conflict events in each country. Variables include measures of "scope," indexed by number of man-days of participation per 100,000 population, and "intensity," indexed by number of deaths per 10 million population. These measures are given for various categories of conflict such as

turmoil; rebellion; and domestic, antforeign, violent and nonviolent conflict. One file contains data on a yearly basis and another consists of 5-year aggregates. CLASS I.

(10) Conflict and Society (ICPSR 7452).

(11) Economic Interpretations of American Intervention, 1964-1968 (ICPSR 7382).

(12) Personality Disorganization Among Refugees of Violence in Columbia, South America (ICPSR 7055).

(13) Contentious Gatherings in Britain, 1758-1834 (ICPSR 8872).

(14) Domestic Violence, International Behavior and National Attributes of Developing Nations, 1962-1967 (ICPSR 5013).

(15) Anti-Foreign Demonstrations in Asia, 1945-1980 (ICPSR 9338).

(16) Political Violence in the United States, 1819-1968 (ICPSR 0080).

(17) Study of Political Violence Attitudes, Personal Experiences With Violence, Emotional Reactions to Assassination and Violence in the Media, 1968 (ICPSR 7354).

(18) Argentina Domestic Violence and Economic Data, 1955-1972 (ICPSR 5213).

(19) World Handbook of Political and Social Indicators II, 1948-1967: Annual Event Data (ICPSR 5028).

(20) World Handbook of Political and Social Indicators II, 1948-1967: Daily Event Data (ICPSR 5215).

(21) Disturbances in France, 1830-1860 and 1930-1960: Intensive Sample (ICPSR 0051).

(22) Strikes and Labor Activities in France, 1830-1960 (ICPSR 8421).

(23) Analysis of Arrests in Paris, June 1848 (ICPSR 0049).

(24) Violent Events in France, 1830-1860 and 1930-1960 (ICPSR 9080).

b. ICPSR Data Bases Listed Under Category III (Conflict, Aggression, Violence, Wars), Sub-category B (Conflict Between and Among Nations).

(1) International Conflict Episodes, 1945-1979 (ICPSR 8303).

Principal Investigator. Alker, Hayward R., and Frank L. Sherman.

Summary. The dynamics of international conflict events are outlined in this collection, with a focus on sequences of events and the roles played by conflict management agents. Every conflict episode is divided into a number of phases—each designating a level of conflict—with a separate record for each phase of each case. The possible levels of conflict range from a quarrel or disagreement claimed by one party to be of international significance (phase 1: dispute) to military action (phase 3: hostilities)

to a resolution of the conflict (phase 6: settlement). Also emphasized are referrals by conflict actors, i.e., initiatives to bring a case to the attention of a potential conflict manager. The collection is divided into five subgroups: (1) a case overview, providing for each case such data as the number of phases and referrals, the presence of management agents, and probable outcome had there been no manager, (2) phase-specific information, including for each phase data on dates, interests at stake, actions taken, alliances, and extent of damages, (3) referral data, including agents involved and support for or opposition to each initiative, (4) data on conflict management agents, detailing the identity, degree of bias, and primary role of each agent, and (5) management agent actions, listing the types of action taken by each agent. CLASS IV.

Extent of Collection. 13 data files.

Data Format. Logical record length.

Parts 1-2: Case Overview. Rectangular file structure, 307 cases, 4 and 13 variables per part, record length 160, 1 record per case.

Parts 3-7: Phase-specific. Rectangular file structure, 929 cases, 11 to 67 variables per part, record length 160, 1 record per case.

Parts 8-10: Management Referral Overview. Rectangular file structure, 491 cases, 5 to 15 variables per part, record length 160, 1 record per case.

Part 11: Conflict Management Agents. Rectangular file structure, 662 cases, 17 variables, record length 160, 1 record per case.

Part 12: Agent Actions. Rectangular file structure, 662 cases, 27 variables per case, record length 160, 1 record per case.

Part 13: Nationality Codes. Rectangular file structure, 542 records, 2 variables per record, record length 160, 1 record per case.

Related Publication. Alker, Hayward R, Jr., and Frank L. Sherman, "Collective Security Seeking Practices Since 1945," in Daniel Frei, ed., *Managing International Crises*, Sage Publications, Beverly Hills, CA 1982.

(2) Conflict and Peace Data Bank (COPDAB), 1948-1978 (ICPSR 7767).

Principal Investigator. Azar, Edward E.

Summary. The Conflict and Peace Data Bank (COPDAB) is a longitudinal computer-based library of daily international and domestic events or interactions. The event records in this file describe the actions of approximately 135 countries in the world, both toward one another and within their domestic environments. A typical descriptive event record, such as international border clash or domestic press censorship, is coded in nine variables. A COPDAB event record includes date of event, actor initiating the event, target of the event, issue area(s), and textual information about the activity. The event record also contains an evaluation by the coder regarding the type and scale value of the event. CLASS IV.

Extent of Collection. One data file plus machine-readable documentation (text). (Documentation is machine-readable only.)

Extent of Processing. BLANKS.

Data Format. Card image. Rectangular file structure, 431,263 cases, 9 variables per case, record length 80 characters, 1 record per case.

(3) Dimensions of Interaction, 1948-1973 (ICPSR 7426).

Principal Investigator. Azar, Edward E., and Thomas J. Sloan.

Summary. Data for this study consists of four data files which are event summaries derived from the Conflict and Peace Data Bank (COPDAB) Project. Machine-readable documentation is also available.

File 1 contains yearly summaries of events directed by one international actor toward another. There are both conflict and cooperation summaries, including the frequency of events, the mean intensity of events in a given year, and a measure of interaction, which combines frequency and intensity. Event summaries are included only for dyads with one of the following political entities as actor and target: Algeria, Canada, Cyprus, Federal Republic of Germany, German Democratic Republic, Egypt, France, Greece, India, Indonesia, Iran, Iraq, Israel, Italy, Japan, Jordan, Kuwait, Lebanon, Libya, Morocco, Pakistan, Palestine Liberation Organization, People's Republic of China, Saudi Arabia, Sudan, Syria, Tunisia, Turkey, United Kingdom, USA, and USSR. The data are recorded for each dyad for each year, 1948-1973. There are 24,180 cases (31 actors by 30 targets for 26 years).

The second file contains domestic event summaries for the same 31 political entities. The variables measure frequency, range, mean, intensity, and dimension of interaction (frequency times intensity) for both conflictive and cooperative domestic events. The data are recorded by year for each entity, a total of 806 cases (31 entities by 26 years).

The third file is similar to the second, but the unit is the dyad. There are 961 cases, one for each actor-target dyad including cases of an entity as both actor and target (31 by 31). One variable is the total number of events initiated by an actor toward the target over the 26-year time period. The second variable is a percentage of the number of events directed at one target relative to all other targets. CLASS II.

Related Publication. Azar, Edward E., and Thomas J. Sloan, *Dimensions of Interaction: A Source Book for the Study of the Behavior of 51 Nations From 1948 to 1973*, Pittsburgh International Association, 1975.

(4) Middle East Event/Interaction Data, 1949-1969 (ICPSR 5201).

Principal Investigator. Blechman, Barry.

Summary. Data on approximately 10,000 events between July 1, 1949 and June 30, 1969 are presented in this data set. Only events in which Israel, Egypt, Syria, Jordan, Lebanon, and Iraq were

actors and targets are included (although third parties acting as mediators are also included as targets). Events are coded for actor, target, the 22 action categories of the World Event/Interaction Survey, arena, and reliability. Source was *The New York Times Index*. CLASS IV.

(5) Middle East Military Event Data, 1949-1969 (ICPSR 5202).

Principal Investigator. Blechman, Barry.

Summary. Data presented in this study document approximately 3,800 events which are strictly physical conflict interactions. The data are described in greater detail than in the study immediately preceding (ICPSR 5201) and have been obtained from a wide range of sources including *The New York Times*, the *Jerusalem Post*, transcripts of UN Security Council meetings and other UN sources. Data consist of date, participants, reports, type of event, and casualties. Data are aggregated to 10-day summaries for each Israel-bordering Arab state dyad. CLASS IV.

Related Publication. Blechman, Barry, "A Quantitative Description of Arab-Israeli Interactions, 1949-1969: Data Sets and Processor," *Professional Paper 78*, Center for Naval Analyses, Washington, DC, September 1971.

(6) CASCON Project: Local Conflict Data, 1945-1969 (ICPSR 5301).

Principal Investigator. Bloomfield, Lincoln, and Robert R. Beattie.

Summary. For 52 local conflicts since 1945, over 500 "factors" are coded for each conflict. "Factors" are conditions or situations which might influence the course of a local conflict toward or away from increased violence. Each factor was recorded as either having no information available, not present, present but no influence, much influence toward violence, some influence toward violence, little influence toward violence, much influence away from violence, some influence away for violence, or little influence away from violence. Factors are grouped into several categories: previous relations between sides, great power involvement, external relations, military strategic, international organizations, ethnic minorities, economic, internal political, characteristic of one side, communication, actions or controls in disputed area. CLASS IV.

Related Publication. Bloomfield, Lincoln, and Robert R. Beattie, "Computers and Policy Making: The CASCON Experiment," *Journal of Conflict Resolution*, Vol XIV, No 4, March 1971.

(7) International Crisis Behavior Project, 1918-1988 (ICPSR 9286).

Principal Investigator. Brecher, Michael and Jonathan Wilkenfeld.

Summary. This data collection was produced as part of the International Crisis Behavior Project, a research effort aimed at investigating 20th-century interstate crises and the behavior of states under externally generated stress. To this end, the data describe, over a 71-year period, the sources, processes, and outcomes of all military-security crises involving states. Variables were collected at both the micro/state actor level and the macro/international level. At the macro level, seven dimensions of crises were measured: crisis setting, crisis breakpoint-exit point, crisis management technique, great

power/superpower activity, international organization involvement, crisis outcome, and crisis severity. Additional macro level variables indicate various aspects of geography, polarity, system level, conflict type, power discrepancy, and involvement by powers. At the state actor level, variables measuring five dimensions of crisis were compiled: crisis trigger, state actor behavior, great power/superpower activity, international organization involvement, and crisis outcome. Additional micro level variables indicate the role of war in each crisis. Others measure several kinds of state attributes: age, territory, regime characteristics, state capability, state values, and social, economic, and political conditions. CLASS IV.

Universe. Part 1: All international crises occurring between December 22, 1917, and December 31, 1988, characterized by the following two conditions: (1) a distortion in the type and an increase in the intensity of disruptive interactions between two or more adversaries, with an accompanying high probability of military hostilities, or, during a war, an adverse change in the military balance, and (2) a challenge to the existing structure of an international system (global, dominant, or subsystem) posed by the higher-than-normal conflictual interactions. Part 2: All foreign policy crises experienced by states due to their involvement in the international crises defined above. The principal investigators define a foreign policy crisis as a situation in which three conditions, deriving from a change in a state's external or internal environment, are perceived by the highest level decisionmakers of the state: (1) a threat to basic values, (2) an awareness of finite time for response to the external threat to basic values, and (3) a high probability of involvement in military hostilities.

Note. The two parts of this collection are linked by common identification numbers. The unit of observation for Part 1, which contains the macro level data, is the international crisis. The unit of observation for Part 2, which contains the micro level data, is the foreign policy crisis experienced by a state. In all, the data cover 390 international crises involving 826 state sectors.

Extent of Collection. Two data files plus machine-readable documentation (text).

Extent of Processing. NONNUM.

Data Format. Card image.

Part 1. ICB1: International Crises. Rectangular file structure, 390 cases, 64 variables, record length 80 characters, 2 records per case.

Part 2. ICB2: Foreign Policy Crises. Rectangular file structure, 826 cases, 85 variables, record length 80 characters, 4 records per case.

Part 3. Code book for all parts. Record length 85 characters. (Documentation is machine-readable only.)

Related Publication. (1) Brecher, Michael, Jonathan Wilkenfeld, and Sheila Moser, *Crises in the Twentieth Century: Vol I--Handbook of International Crises*. Oxford: Pergamon, 1988. (2) Brecher, Michael, Jonathan Wilkenfeld, and Sheila Moser, *Crises in the Twentieth Century: Vol II--Handbook*

of *Foreign Policy Crises*. Oxford: Pergamon, 1988. (3) Brecher, Michael, Jonathan Wilkenfeld, and Sheila Moser, *Crises in the Twentieth Century: Vol III—Crisis, Conflict, and War*. Oxford: Pergamon, 1988.

(8) Interstate Security Conflicts, 1945-1974 (ICPSR 7536).

Principal Investigator. Butterworth, Robert L.

Summary. This study contains data on 310 cases of interstate conflict which had a political-security focus and occurred between 1945 and 1974. There are 47 variables divided into four categories: case identification, conflict characteristics, management action, and management influence. CLASS IV.

Related Publication. Bloomfield, Robert L., *Managing Interstate Conflict, 1945-1974*, University Center for International Studies, University of Pittsburgh, 1976.

(9) US and USSR Crises, 1946-1976 (ICPSR 7702).

Principal Investigator. CACI, Inc.-Federal.

Summary. This study contains three files of information on international crises: two deal with 307 crises that were of concern to the Soviet Union, 1946-1975. The research strategy for the analysis of Soviet crises was based on a decision to describe the crisis events as seen through the eyes of the Soviets. At the same time, major efforts were made to collect data compatible with those used in earlier analyses of US crisis behavior. Therefore, Soviet sources were used to identify the set of USSR crises while it was necessary to supplement data not available in Soviet media sources with information from Western media sources.

Data for the USSR consist of (1) 386 USSR crises and (2) 101 crises occurring between 1956 and 1976 selected from the list of 386 crises to reflect the Soviet policy process, (3) 43 management problems associated with the 101 crises of the previous file, (4) 59 Soviet objectives associated with the 101 crises, and (5) 64 actions associated with the 101 crisis situations; and data including the 101 crises coded for the presence or absence of 59 Soviet objectives, coded for 64 USSR actions, and coded for 43 Soviet management problems, and the full set of 386 crises coded for 29 general characteristics. The two files dealing with crises of concern to the United States were coded using Western media sources. The data are similar to the USSR files; 307 US crises were identified and grouped to form 101 crises; also included in the file are 57 US actions, 48 US objectives, 79 US management problems, and 21 crisis characteristics. CLASS II.

(10) Political Conflicts, 1944-1966 (ICPSR 5302).

Principal Investigator. Cady, Richard and William Prince.

Summary. Data for 323 political conflicts which occurred in the years 1944-1966 were supplied by the Social Science Division, Bendix Corporation. For each conflict, the variables include measures of duration, the type of military operations, the type of conflict, the method of termination, and the outcome with regard to the United States. CLASS II.

(11) Nations in Conflict: Data on National Growth and International Violence for Six European Major Powers, 1870-1914 (ICPSR 7425).

Principal Investigator. Choucri, Nazli, and Robert C. North.

Summary. The data were compiled as part of the investigators' study dealing with the dynamics of conflict and warfare and the role of national growth and expansion in that process. The annual aggregate data compiled for the period 1870-1914 on Britain, France, Germany, Italy, Russia, and Austria-Hungary are in eight categories: national size, colonial size, economic and productivity profile, commercial activity, government budget, alliances, violence, and conflicts of interest.

The first six categories are aggregate data compiled from yearbooks and other historical sources. One of the alliance measures, adversary relationship, is a dummy variable coded "1" if nation was not aligned, "0" if aligned. Variables in the last two categories are derived from subjective scaling procedures. CLASS II.

Related Publication. Choucri, Nazli, and Robert C. North, *Nations in Conflict*, W. H. Freeman and Co., San Francisco, 1975.

(12) Transactional Data Bank of International Conflict and Amity Events (ICPSR 5210).

Principal Investigator. Feierabend, Ivo, Rosalind Feierabend, and J. S. Chambers.

Summary. This data set contains data on international hostile and amity events. Sixteen types of hostile transactional events are presented in this collection, ranging from protests, accusations, and recall of officials to quasi-military actions, troop mobilizations, and war. Fourteen types of amity events are also covered, ranging from offers to negotiate and confer to exchanges, agreements, and alliances. Events are qualified in 19 categories, including date, actor, duration, and persons involved. The direction of the event and its retaliatory or nonretaliatory character are also included. These data are of a preliminary nature. Data on approximately 7,000 international events are included in this collection. CLASS IV.

(13) Conflict Management by International Organizations, 1945-1970 (ICPSR 5303).

Principal Investigator. Haas, Ernst, Robert Butterworth, and Joseph Nye.

Summary. This data set contains 25 variables for 132 disputes in which the authors found evidence of some involvement by the United Nations, the Organization of American States, the Organization of African Unity, the Arab League, or the Council of Europe. Variables measure aspects of the dispute including the world political conditions, intensity, spread, system era, power of parties, issues, and organizational response of the alignment which includes the leadership, consensus, forum and operations, and the degree of success or failure of the organization. CLASS II.

Related Publication. Haas, Ernst, *et al.*, *Conflict Management by International Organizations*, Morristown, NJ, General Learning Press, 1972.

(14) **Behavioral Correlates of War, 1816-1975 (ICPSR 8606).**

Principal Investigator. Leng, Russell J.

Summary. The behavior of various states involved in 38 interstate crises over the past 160 years is the focus of this data collection. More specifically, these data follow the attempts of these states to influence others and the responses generated over the course of the crises. The sample reflects a broad historical range of crises and includes those that involve war as well as those that do not. Each data file contains information on a crisis or set of crises. Within each data file, the cases represent actions taken during the crisis period. The data allow for a micro level of analysis with measures that distinguish a wide range of cooperative and conflictive interstate actions. In addition, measures are included that are sensitive to the mix of different behaviors and that record the tempo of the action within given time intervals. The Crisis program provided allows for the counting and scaling of events. CLASS IV.

Universe. Identified population of interstate crises, 1816-1975.

Sampling. Stratified sample of interstate crises.

Note. (1) Part 32 is the Crisis program, an interactive program written in VAX-11 PASCAL, which contains subroutines for counting and scaling actions. This program was written for use on a Digital VAX minicomputer. Part 33 is a categorization and weight scheme used by the Crisis program. This scheme is necessary to run the Crisis program. It may be modified by users to their own specifications. It also should be noted that, in order to use the Crisis program, the names of the crisis files and of the categorization/weight scheme file *must* match the part names provided below, excluding the section in parentheses. (2) Each part contains 44 variables, with logical record lengths of 80 characters. There is one record per case. All data files are rectangular. There are 88 to 2,352 cases per part in this collection.

Extent of Collection. Thirty-one data files plus two computer program files.

Data Format. Card image.

Part 1: Pastry.dat (Pastry War).

Part 2: Schles.dat (2d Schleswig-Holstein War).

Part 3: Rustrk.dat (Russo-Turkish War and Britain-Russian Crisis).

Part 4: Brprt.dat (British-Portugal Crisis).

Part 5: Spanm.dat (Spanish-American War).

Part 6: Fashod.dat (Fashoda Crisis).

Part 7: 1stmor.dat (1st Moroccan Crisis).

Part 8: Cenam.dat (2d Central American War).

Part 9: Bosnia.d (Bosnia Crisis).

Part 10: 2dmor.dat (2d Moroccan Crisis).

Part 11: Balkan.dat (1st and 2d Balkan War).

Part 12: Preww1.dat (Pre-World War I).

- Part 13: Teshen.dat (Tenschen Crisis).
- Part 14: Chaco.dat (Chaco Dispute).
- Part 15: Italet.dat (Italo-Ethiopian War).
- Part 16: Rhine.dat (Rhineland Crisis)
- Part 17: Ansch.dat (Anschulss Crisis).
- Part 18: Munich.dat (Munich Crisis).
- Part 19: Pollit.dat (Polish-Lithuanian Crisis).
- Part 20: Preww2.dat (Pre-World War II).
- Part 21: Kash1.dat (1st Kashmir War).
- Part 22: Palest.dat (Palestine War).
- Part 23: Berair.dat (Berlin Blockade).
- Part 24: Triest.dat (Trieste Crisis).
- Part 25: Suez.dat (Suez Crisis).
- Part 26: Berwal.dat (Berlin Wall Crisis).
- Part 27: Cuban.dat (Cuban Missile Crisis).
- Part 28: Cyprus.dat (Cyprus Crisis).
- Part 29: Kash2.dat (2d Kashmir War and Rann of Kutch).
- Part 30: Sixday.dat (Six-Day War).
- Part 31: Bangla.dat (Bangladesh War).
- Part 32: Crisis.pas (Crisis program).
- Part 33: Both.sco (Categorization and weighting scheme).

Related Publication. (1) Leng, Russell, J., "Reagan and the Russians: Crisis Bargaining Beliefs and the Historical Record," *American Political Science Review* 78 (June 1984), pp 338-355. (2) Leng, Russell J., and Charles Gochman, "Dangerous Disputes: A Study of Conflict Behavior and War," *American Journal of Political Science* 26 (November 1982), pp 664-687.

(15) Decisions, Events, and Perceptions in International Crises: The Arab-Israel Conflict, 1967-1973 (ICPSR 5307).

Principal Investigator. McCormick, David.

Summary. The data are daily aggregations for the 1967 Arab-Israeli conflict, May 1 to July 31, and the 1973 Arab-Israeli conflict, September 1 to December 31, for three general sets of variables: (1) manual content analyses of the public statements of Israeli leaders, (2) computer content analyses of the public statements of Israeli leaders, and (3) Arab → Israeli events and Israeli → Arab events. The manual content analyses contain a measure of the perception of time pressure, the components of threat perception, and indices of threat perception. The computer content analysis data include the Osgood semantic differential measures and a combined index of threat perception. The event data were obtained from World Event/Interaction Survey (WEIS), M. Mansoor, *Jerusalem Post*, and other

scattered sources. They were coded using WEIS categories, and Walter Corson's and Edward Azar's indexing systems. Data were obtained from the First Ann Arbor Corporation. CLASS IV.

(16) International Terrorism: Attributes of Terrorist Events, 1968-1977 (ICPSR 7947).

Principal Investigator. Mickolus, Edward F.

Summary. This four-part study substantially expands the coverage of terrorist incidents first reported in the earlier data collection, "International Terrorism: Attributes of Terrorist Events (ITERATE)," which covered January 1970 through July 1973. This data set contains data on 3,329 international terrorist attacks from 1968 through 1977. The Common File includes information on the type of attack, the location of the beginning and end of the incident, the name of the terrorist group involved, and the numbers of deaths and injuries. The Hostage and Fate Files provide more detailed information on the characteristics and fates of both the hostages and the terrorists. The last file consists of data on skyjacking incidents. All three of the supplementary files can be linked to the Common File when appropriate.

The Common File has 37 variables and 3,329 cases. The Hostage File has 52 variables and 315 cases; the Fate File, 19 variables and 415 cases; and the Skyjack File has 24 variables and 617 cases. CLASS I.

(17) War Ledger Data, 1870-19. (ICPSR 9000).

Principal Investigator. Organski, A. F. K., and Jacek Kugler.

Summary. The five data sets comprising this study examine the relationship between national power and war. In the first data set, the conditions preceding the outbreak of war are examined; in particular, the links between national growth, the distribution of power, and the onset of war. The second data set contains variables which measure national capabilities and estimate political capacities. The third data set relates to the consequences of war in terms of national power and growth. The final two data sets concern nuclear deterrence. The fourth data set contains crisis data to test whether deterrence works and the last data set contains information on US and USSR defense expenditures in order to establish if the two major nuclear powers are in a nuclear arms race. CLASS III.

Universe. Part 1: International conflicts where major powers actively participated. Part 2: All nations involved in international conflict. Part 3: Nations involved in World Wars I and II. Part 4: All crisis events from 1945 to 1974. Part 5: US and USSR defense spending and resources, 1945-1978.

Extent of Collection. Five data files.

Data Format. Logical record length.

Part 1. Causes of war. Rectangular file structure, 126 cases, 8 variables, record length 15, 1 record per case.

Part 2. National capabilities. Rectangular file structure, 988 cases, 10 variables, record length 43, 1 record per case.

Part 3. GNP index. Rectangular file structure, 101 cases, 22 variables, record length 87, 1 record per case.

Part 4. Crisis events data. Rectangular file structure, 366 cases, 6 variables, record length 8, 1 record per case.

Part 5. US/USSR arms race. Rectangular file structure, 32 cases, 24 variables, record length 98, 1 record per case.

Related Publication. Organski, A. F. K., and Jacek Kugler, *The War Ledger*, University of Chicago Press, Chicago, 1980.

(18) International Military Intervention, 1946-1988 (ICPSR 6035).

Principal Investigator. Pearson, Frederic S., and Robert A. Baumann.

Summary. This data collection documents all cases of military intervention across international boundaries by regular armed forces of independent states in the regions of Europe, the Americas (and Caribbean), Asia and the Pacific, Sub-Saharan Africa, and the Middle East/North Africa. Military interventions are defined operationally in this collection as the movement of regular troops or forces (airborne, seaborne, shelling, etc.) of one country into the territory or territorial waters of another country, or forceful military action by troops already stationed by one country inside another, in the context of some political issue or dispute. The study seeks to identify politically important actions which interpose a state directly into the conflict patterns occurring in another state, and which conceivably involve a breach of the sovereignty of the target state (albeit by invitation in some cases). The collection identifies the intervener and target countries and specifies the starting and ending dates of the intervention. A series of potential interests in or motives for intervention are presented, including effects of the target's domestic disputes, foreign or domestic policies, and efforts to protect social factions in the target, to attack rebels in sanctuaries across borders ("hot pursuit"), to protect or enhance economic/resource interests, to protect military or diplomatic facilities, to save lives, or to affect regional power balances and strategic relations. Information is provided on the direction of the intervention, i.e., to support or oppose the target government, to support or oppose opposition groups in the target, or to support or oppose third party government or opposition groups. Other variables show the degree of prior intervention, the alliance or treaty relationship between intervener and target, prior colonial status, prior intervention, and measures of intervener and target power size. A series of intensity measures, such as battle-related casualties, is also included. For each type of incursion, by land, sea, or air, an ordinal scale of involvement is presented, ranging from minor engagement such as evacuation, to patrols, acts of intimidation, and actual firing, shelling, or bombing. Finally, contiguity information is provided to indicate both whether intervener and target are geographically contiguous, and whether the intervention was launched from contiguous territory. CLASS III.

Universe. All cases of military interventions from 1946 through 1988.

Note. Part 2 of this collection contains SAS language statements, a data list, instream data, and other program statements to read the files directly into SAS.

Extent of Collection. One data file, plus machine-readable documentation (text), and an accompanying computer program.

Extent of Processing. NONNUM/BLANKS/MDATA.

Data Format. Logical record length.

Part 1. Main data file. Rectangular file structure, 667 cases, 35 variables, record length 134 characters, 1 record per case.

Part 2. SAS program file with instream data. Record length 80 characters.

Related Publication. (1) Pearson, Frederic S., "US-Soviet Competitive Intervention: Retrospect and Prospect," in Manuel J. Pelaez (ed) *Public Law and Comparative Politics. Trabajos en Homenaje a Ferran Vallis i Taberner*, Vol XVII, Barcelona, Spain; Facultad de Derecho de la Universidad de Malaga, et al., 1991, pp 4985-5017. (2) Pearson, Frederic S., and Robert A. Baumann, "International Military Intervention in Sub-Saharan African Subsystems," *Journal of Political and Military Sociology* 17 (Spring 1989), pp 115-150. (3) Pearson, Frederic S., Robert A. Baumann, and Gordon N. Bardos, "Arms Transfers: Effects on African Interstate Wars and Interventions," *Conflict Quarterly* (Winter 1989), pp 36-62.

(19) Statistics of Deadly Quarrels, 1809-1949 (ICPSR 5407).

Principal Investigator. Richardson, Lewis Fry.

Summary. Data for 779 dyadic quarrels from some 300 conflicts were supplied by Rudolph Rummel and cover the time period from 1809 to 1949. A dyadic quarrel is a situation involving a pair of opponents and resulting in more than 315 human deaths. The magnitude of a quarrel is measured by the logarithm of the number of deaths. The range of magnitude in the study is from 2.5 to 7.5, the latter figure for nations involved in World War II. Each quarrel is identified by its beginning date and magnitude. For each quarrel, the nominal variables include the type of quarrel, as well as political, cultural, and economic similarities and dissimilarities between the pair of combatants. CLASS II.

Related Publication. Richardson, Lewis Fry, *The Statistics of Deadly Quarrels*, Quadrangle Press, Chicago, 1960.

(20) Dimensionality of Nations Project: Dyadic Foreign Conflict Variables, 1950-1965 (ICPSR 5408).

Principal Investigator. Rummel, Rudolph J.

Summary. This study contains data on 1,557 nation dyads, including any nation with conflict behavior in the years 1950, 1955, 1960, 1963, or 1965. There are 19 characteristics and conflict behavior variables. There are approximately 1,557 card-image equivalents. CLASS II.

(21) Foreign Conflict Behavior, 1950-1968 (ICPSR 5214).

Principal Investigator. Rummel, Rudolph J.

Summary. This data set contains approximately 13,000 events such as border clashes and threats on over 30 descriptive variables for 82 nations. The source of the data as well as measures of its reliability have also been coded. The periods of time covered include 1950, 1955, 1960, and 1962-1968. CLASS IV.

(22) Dimensions of Conflict Behavior Within and Between Nations, 1955-1960 (ICPSR 5021).

Principal Investigator. Rummel, Rudolph J., and Raymond Tanter.

Summary. This data collection, containing 22 variables for 86 countries, was the product of two separate studies conducted by Rudolph J. Rummel and Raymond Tanter utilizing identical variables for the time periods 1955-1957 and 1958-1960, respectively. The variables describe domestic conflict behavior such as riots and coups, and foreign conflict behavior such as protests and threats. CLASS I.

Related Publication. (1) Rummel, Rudolph J., "Dimensions of Conflict Behavior Within and Between Nations," (2) Tanter, Raymond, "Dimensions of Conflict Behavior Within and Between Nations, 1958-1960," both in John Gillespie and Betty Nasvold (eds) *Macro-Quantitative Analysis*, Sage Publications, Beverly Hills, CA, 1971.

(23) Wages of War, 1816-1980: Augmented With Disputes and Civil War Data (ICPSR 9044).

Principal Investigator. Singer, J. David, and Melvin Small.

Summary. This data collection consists of three separate data sets. Data were collected to study the trends and changes in the frequency magnitude, severity, and intensity of international wars, civil wars, and international disputes. For each data set, the unit of analysis is the participant in a particular conflict. While the three data sets are related, they are mutually exclusive in that each describes a particular type of war (interstate or civil) or a dispute.

The first collection describes the experience of each interstate system member in each war. The file contains 302 records, each containing 42 variables that describe those experiences. In order to be considered a nation a participant, certain minimal criteria of population and diplomatic recognition were used. Qualifying nation participants are classified as to whether or not they were members of the European central system at the time of the war and, therefore, active and influential in European centered diplomacy. The geographical location of the war is coded as well as the severity of the war as determined by its duration and the number of battle-connected deaths. The prewar population of each nation participant is also coded.

The second data set is a study of 106 major civil wars involving 139 participants between 1816 and 1980. This file contains 139 records, each containing 36 variables. An internal war is classified as a major civil war if (a) military action was involved, (b) the national government at the time was

actively involved, (c) effective resistance (as measured by the ratio of fatalities of the weaker to the stronger forces) occurred on both sides and (d) at least 1,000 battle deaths resulted during the civil war. The geographical area in which the war was fought is also coded as well as whether nations outside the civil war actively and overtly participated on one side or the other. The duration, beginning, and ending dates of the civil war, and the prewar population and number in the armed forces of each participant are also included.

The third data set describes the international disputes that occurred from 1817 through 1967. International disputes are distinguished from international wars if fewer than 1,000 total fatalities are incurred by all the disputants involved. The file contains 967 records, each containing 51 variables. The participants involved in each dispute are identified using Singer-Small nation numbers. The level of hostility reached by each participant is also coded as well as whether or not a major power was among the nations involved in the dispute, the beginning and ending dates of each dispute, the total number of noncivilian fatalities that resulted, and any special characteristics that may be associated with a dispute (e.g., whether a dispute eventually escalated into a Singer-Small war). CLASS II.

Note. As of September 1993, the third data set on international disputes between 1817 and 1976 was unavailable awaiting an update from the principal investigators.

Related Publication. (1) Singer, J. David, and Small, Melvin, *The Wages of War: A Statistical Handbook*, John Wiley & Sons, Inc., New York, ISBN: 0-471-79300-0. 1972, 419 pp, UNCLASSIFIED, available from libraries. (2) Small, Melvin, and Singer, J. David, *Resort to Arms: International and Civil Wars, 1816-1980*, Sage Publications, Beverly Hills, 1982, ISBN: 0-8039-1776-7, 1982, 372 pp, UNCLASSIFIED, available from publisher and libraries. (3) Singer, J. David (ed), *Correlates of War*, in two volumes, The Free Press (a division of Macmillan Publishing Co., NY); Vol I, 1979, ISBN: 0-02-928960-2, UNCLASSIFIED, available from libraries; Vol II, 1980, ISBN: 0-02-92010-4, UNCLASSIFIED, available from libraries (CAA Lib. No. = 327.11 COR).

(24) *World Handbook of Political and Social Indicators II, 1948-1967: Intervention Data* (ICPSR 5306).

Principal Investigator. Taylor, Charles L., Michael C. Hudson, and John D. Sullivan.

Summary. This data set contains data for interventions recorded at daily intervals during the 20-year period 1948-1967. The daily report is the unit of analysis. There are 1,073 records, one for each day on which an intervention occurred in a country. The number of records per country varies. If, for example, a country had no action meeting the criteria for inclusion as an intervention, no record is given for that day for that country. Those countries not involved in an intervention were excluded. Data are recorded for 89 of the 136 nations in the *World Handbook's* Aggregate Data file and 2 international organizations. Some of the 31 variables included are the number of interveners, type of group involved, air and naval incursions, and length of intervener's presence in the country. The data

sources are *The New York Times Index*, Associated Press, *Asian Recorder*, *African Research Bulletin*, *Middle East Journal*, and *African Diary*. CLASS I.

Related Publication. Taylor, Charles L. and Michael C. Hudson, *World Handbook of Political and Social Indicators*, Second Edition, Yale University Press, New Haven, CT, 1971.

(25) Study of War: Warlikeness and Other Characteristics of Primitive Peoples (ICPSR 5905).

Principal Investigator. Wright, Quincy.

Summary. Data were compiled on 652 primitive peoples. The data were keypunched directly from *A Study of War*, Appendix IX, Table 5. Alphabetic codes were changed to numeric ones. Variables include religion, character of war, habitat, race, culture, political, and social organization, and intercultural relations of the primitive peoples. CLASS IV.

Related Publication. Wright, Quincy, *A Study of War*, Univ. of Chicago Press, Chicago, 1940.

(THIS PAGE INTENTIONALLY LEFT BLANK)

APPENDIX D

OTHER DATA BASES OR SOURCES OF INFORMATION

D-1. INTRODUCTION. This appendix describes several other data bases that deserve to be mentioned in this paper, even if for some reason they do not satisfy the criteria for inclusion in Appendix B's catalog of data bases. For example, a data base might be proprietary or otherwise require special handling; or it may not be relevant to personnel attrition rates in land combat operations (though providing other information of interest to military operations analysts); or it may not have been reducible to electronic form within the level of effort available to the PAR studies.

Some examples of proprietary data bases are those maintained and provided to researchers by the Inter-University Consortium for Political and Social Research. This organization's goals and procedures are described briefly in Appendix C, together with some of its data bases that seem most likely to be of interest to military operations analysts.

This appendix describes some additional data bases or potential sources of information. They are listed here because they may be of interest to investigators somewhat outside the intended primary target audience of this paper (namely, that part of the US Army military operations research group community involved in studying personnel attrition for the purposes of forecasting its future levels, understanding its dynamics and implications, or of analyzing what effects different weapons and/or protective gear have upon personnel attrition). There is certainly no attempt to make the following a comprehensive list—it mentions only those items that accidentally came to our attention while compiling the catalog of data bases available on diskette.

D-2. OTHER DATA BASES

a. **Anti-Armor Defense Data (A2D2).** This study was conducted at CAA in the period Jan 88–Feb 91. It built a data base describing quantitatively the forces and actions involved in several antiarmor battles involving US forces, for use in determining combat degradation factors for armor combat. The aim was to collect detailed historical data on at least five combat actions where US forces were defending against enemy armor attacks. The data set will be used in a joint US/UK analysis of the degradation in antiarmor defense effectiveness under combat conditions. The resulting degradation factors will allow projections of the combat performance of future antiarmor defenses to be based upon a balanced combination of historical and instrumented field test data. Its products were

(1) Science Applications International Corporation Draft Final Technical Report, "Anti-Armor Defense Data Study (A2D2), Draft Final Report, Volume I—Technical Report," 9 March 1990.

(2) Science Applications International Corporation Report 90-1174, "Anti-Armor Defense Data Study (A2D2), Phase I, Final Report, Volume II--US Anti-Tank Defense at Mortain, France (August, 1944)," 30 March 1990.

(3) Science Applications International Corporation Report 90-1306, "Anti-Armor Defense Data Study (A2D2), Phase I, Draft Final Report, Volume III--US Anti-Tank Defense at Dom Bütgenbach, Belgium (December, 1944)," 10 September 1990.

(4) Science Applications International Corporation Report 91-1012, "How to Research' Guide," 15 February 1991.

(5) Science Applications International Corporation Report 91-1012, "Anti-Armor Defense Data Study (A2D2), Phase II, Final Report, Volume IV--US Anti-Tank Defense at Krinkelt-Rocherath, Belgium (December 1944)," 15 February 1991.

(6) Science Applications International Corporation, "Action Chronology: Krinkelt-Rocherath, Belgium, 16-20 December 1944," 15 February 1991.

This project has since been taken over by the United Kingdom's Defence Operational Analysis Centre, Brooklands, Parvis Road, West Byfleet, Surrey, KT14 6LY, United Kingdom. They are the proper point of contact for further information on this project's current status.

b. Evolution of the US Army Force Structure (EUSAFS). This study was conducted during the period Sep 88--Jul 89. It provided a convenient, readily accessible source of detailed factual information on the evolution of US Army force structure from 1910 through the late 1980s. It aimed at clearly delineating the main features of US Army force structure evolution since 1910. This was motivated by the recognition that no convenient source of such information was then available. The scope of the work includes partitioning (*i.e.*, allocating into exhaustive and mutually exclusive categories) all US Army and host nation support personnel by: sex, branch of service, unit assignment, geographical deployment, envisioned area of commitment, component, major command assignment, major force category, rank, and standard requirement code (SRC). The Essential Elements of Analysis include such items as:

(1) What definitions, measures, or equivalences need to be developed to allow consistent comparisons of force structures, components, and units over the last 80 years?

(2) What kinds of data are really needed to understand the evolution of force structure and to derive lessons learned for evaluating current and future force structures?

(3) Which of these data are not now readily available, and what is apparently required to assemble those data?

(4) How has combat service support structure varied over time?

(5) What can one conclude from available data about the effect of forward deployment and modern strategic lift capability on force structure?

(6) How did the advent of modern air power apparently influence the evolution of Army force structure?

(7) How did the advent of nuclear warfare affect Army force structure development?

(8) What do the data say about the influence of the development of armored combat vehicles and antiarmor weapons on force structure?

(9) What is the evidence of a relationship between Active, Reserve, National Guard, and host nation support evolution?

(10) What do the data suggest about trends in the relationship between general purpose forces and general support forces over time?

(11) What force structure changes have accompanied force modernization efforts?

Envisioned applications include (i) suggesting force structure issues for theater-level gaming and analysis, (ii) comparing force structure requirements determinations with historical force structure trends and evolution, (iii) providing facts essential to a deeper understanding of the reasons behind force structure changes. Important issues, topics for future research, and lessons learned are implicit in the results.

The products of this work are reported in Final Report (Vol I and II), SAIC Report 89-1495, "Evolution of US Army Force Structure (EUSAFS), July 1989, AD-B-135-297L (Vol 1), AD-B-135-297L (Vol 2).

c. Simple Combat Attrition Law Evaluation Data (SCALED). This work was conducted mainly during the period Apr 89-Mar 91. It was to provide reliably complete and accurate historical combat data for use in evaluating empirically a variety of simple attrition laws. The specific objective was to (i) gather, (ii) compile and organize into systematic data files, and (iii) document detailed historical data on the strengths, gains, and losses of the engaged forces in at least 12 battles. This was motivated by a recognition that the currently available data of this type are seriously inadequate, and a recognition that more complete data are essential for model validation efforts. The scope of the work included selecting suitable candidate battles; devising a satisfactory data collection plan; collecting, compiling, organizing and appraising the data; identifying the forces engaged in the battle; summarizing the temporal evolution of their strengths, gains and losses; preparing a narrative account of the action; and documenting the results in a technical report supported by systematically organized data files. The following battles were selected for detailed examination: Antietam, Belleau Wood

(World War I), Gettysburg, Second Bull Run, Metz Campaign (Franco-Prussian War of 1870), Waterloo, Westwall Battle (World War II). The products of this work are on file at CAA.

d. CODA (Combat Operational Data Analysis). This work was conducted during the period Jun 88-Feb 89. It was aimed at compiling and reexamining data on the suppression effects of artillery in World War II as reported in contemporaneous British Army Operational Research unit reports. The insights obtained from these data were documented, and their applicability assessed. The specific objective was to compile, summarize, and perform some preliminary analyses on data from existing operational analysis reports generated during WWII. This was motivated by the view that a more thorough collection and evaluation of these data can be used to improve current US Army wargames and studies, as well as preserving them for future use. The scope of the work includes extracting data from existing combat operational reports and evaluating their utility (i) as a basis for modeling suppression effects due to fire support and (ii) as a source of benchmarks or reference points against which to judge artillery expenditure results.

The products of this work are reported in Science Applications International Corporation Report SAIC-89/1098, "Combat Operational Data Analysis: An Examination of World War II Suppression Data," with computerized data base, 28 February 1989. It concluded that the data can be used to suggest hypotheses and to identify promising research topics, but that they are not adequate for modeling suppression.

e. Quincy Wright's book *A Study of War*, University of Chicago Press, 1st edition, 1942, 2d edition, 1965, is a classic volume on warfare that contains (primarily in its appendixes) tabulations of population growth, warlike characteristics of primitive peoples, number of battles engaged in by principal European powers (1480-1940), duration of battles by centuries (1618-1905), wars of modern civilization (1480-1941), Sorokin's estimate of war casualties in Europe by centuries (1000-1925), proportion of the population of France and Great Britain that died in war (1600-1930), and many others. However, its data on attrition in battles are covered in more detail by the BERNDT-1993, BWSH-1993, CDB90-1991, and similar data bases. In addition, its data on attrition in wars are covered in more detail by the Singer and Small data bases (see the ICPSR-1984 data base).

f. Smith, E. G.; and Donovan, G. N., "Trends in Warfare," British Army Operational Research Group, Memorandum No. F.6, April 1955, 38 pp, RESTRICTED, available from United Kingdom Defence Operational Analysis Establishment (DOAE). This noteworthy study is very interesting and informative. It is omitted in part because of security restrictions, but mainly for the same reasons cited above for Wright's *Study of War*.

g. Data from Korotkin, I. M., *Battle Damage to Surface Ships During World War II* (BOEVYE POVREZHDENŪA NADOVDNYKH KORABLEĬ), Leningrad, Sudpromgiz, 1960, translated by the US Joint Publications Service Branch for the David Taylor Model Basin, February 1964, Translation 310, S-F013 04 03, Task 1759, UNCLASSIFIED, available from DTIC as AD-437 233. This 467-page translation makes available in English a systematic compendium of 102 cases of battle damage and loss to surface ships of the four basic classes employed in World War II (aircraft carriers, battleships, cruisers, and destroyers), and embraces practically the entire range of displacements (from 1,000 to 70,000 tons) and all the various architectural types of surface ships. The effects of torpedo explosions, mine explosions, and bomb bursts upon ships are analyzed and also the simultaneous effects of projectiles. In addition, some characteristic cases of shelling and of Japanese kamikaze aircraft attack are considered. It does not discuss damage to Soviet ship or Soviet trends in warship design. Personnel losses are often included in the description of damage. This important work was excluded because it did not deal with land combat operations.

In this connection, we note that Mr. Richard L. Humphrey, formerly with the US Naval Surface Warfare Center, has summarized and reanalyzed Korotkin's material and added additional information from selected conflicts that occurred since World War II. He presented some of this work at the Thirteenth General Working Meeting of The Military Conflict Institute, 9-12 October 1992. He may be contacted at 9300 Colesville Road, Silver Spring, MD 20901-4818.

h. Data from the Joint Technical Coordinating Group (JTTCG) for Munitions Effectiveness, "Evaluation of Wound Data and Munitions Effectiveness in Vietnam (WDMEV)," US Department of Defense, Final Report, December 1970, In three volumes (Vol 1 is 406 pp, Vol 2 is 289 pp, Vol 3 is 275 pp), Vols 1 and 3 are UNCLASSIFIED, Vol 2 is SECRET. Available from DTIC (Vol 1 is AD-879 516, Vol 2 is AD-513 342L, Vol 3 is AD-879 517).

The original Wound Data and Munitions Effects Team (WDMET) data consists of detailed descriptions of the wounds for nearly 8,000 US Army and Marine Corps casualties during the period of WDMET operations in Vietnam, i.e., from June 1967 through June 1969. Information on each casualty includes the location of wounds, activity of the wounded person at the time wounds were received, and an assessment of the severity of the wound. Detailed anatomical data are provided regarding the wound tract. A brief synopsis of the action in which the wounds were received is included. The data are based primarily on personal observation and studies by the WDMET of witness and treatment provider interviews, and on surgical, pathological, and autopsy reports. The original data are in the form of 1969 vintage mainframe magnetic tape, audio tape, photographic film, X-rays, and artifacts.

Between 1987 and 1990, the Casualty Care Research Center (CCRC) reentered selected fields of most cases on Macintosh computers as part of their "TRAUMABASE" multimedia data base project. These data currently exist in a 57 megabyte Acius 4th Dimension Data Base on hard disk. Multimedia data were not computerized, except for a test sample of 200 cases—this is believed to be the only existing collection of multimedia data on wounds received in land combat operations conducted by US armed forces.

Diskette versions of the data base are not available. The physical records are confidential patient information currently in the custody of the CCRC, under the Department of Military and Emergency Medicine, Uniformed Services University of the Health Sciences (USUHS), located on the grounds of the Naval Medical Center, 4301 Jones Bridge Road, Bethesda, MD 20814. Access requires authorization granted in accordance with USUHS Procedure 6406.

i. The Multinational Data Base Study, conducted by the Casualty Care Research Center of the Uniformed Services University of the Health Sciences, is an effort to collect data on wounding by tactical environment, causative agent, and early outcomes from all countries willing to share such information. As part of this, they were as of September 1993, very close to an agreement with the Russians to obtain data from the Afghanistan War. This is an intriguing effort to assemble data on a multinational basis, which will facilitate comparisons of certain aspects of the casualty experience of one nation with that of another. For more information on current status, contact the Casualty Care Research Center of the Uniformed Services University of the Health Sciences, located on the grounds of the Naval Medical Center, 4301 Jones Bridge Road, Bethesda, MD 20814. It was omitted because it is a work in progress and therefore is constantly being changed and revised.

j. The Counter Narcotics Tactical Operations Medical Support (CONTOMS) data base is the product of an ongoing accumulation effort. As of January 1994, it included about 1,000 incident reports from a range of law enforcement activities. Over 200 casualties, some with multiple injuries, have been reported. The information includes the nature of the wounds and the circumstances under which they were incurred. The data are currently kept in a Borland Paradox 4.0 data base.

Information on each casualty includes the type of operation; date, time, and duration; initial personnel numbers of Special Weapons and Tactics (SWAT) team members, supporting law enforcement personnel, bystanders, and perpetrators; numbers of casualties from each of the groups; type of wounding agent; and the nature of wounds inflicted. Casualties are identified as wounded or killed. A description of the operational circumstances is included. This data base provides information on nonlethal as well as lethal injuries and medical support information not available from standard crime reports. The information is obtained from reports submitted by law enforcement teams operating as special action groups in US police operations. Although a diskette is not available, the transfer of Paradox files or exportation to other formats is possible.

It is generally conceded that the wounds suffered or inflicted by law enforcement agencies are not in all respects comparable to those received in military combat action. Accordingly, extrapolations from SWAT team experience to military combat situations must be done with great care and with full appreciation of the differences between these two situations. For further information, contact the Casualty Care Research Center of the Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814.

k. A data base of military and civilian deaths in wars, as reported in the annual publication *World Military and Social Expenditures*, edited by Ruth Leger Sivard and published by World Priorities, Box 25140, Washington, DC 20007, UNCLASSIFIED, is available from the publisher and libraries. The 1993 edition is ISSN 0363 4795, ISBN 0-918281-05-9. This is the latest issue of this publication, which has appeared annually since 1978. The main purpose of these reports is to provide an annual accounting for the use of world resources for social and for military purposes, and thereby to furnish an objective basis for assessing relative priorities. In bringing together military costs and social needs for direct comparison, the report bridges a gap in the information otherwise available to the public. Each annual issue of this publication addresses primarily the course of expenditures over the last 10 to 20 years. However, the material on wars and war-related deaths provided in the 1988 issue covers the period from 1700 to about 1988. The 1993 issue extends the latter date to 1993, but omits the values for dates earlier than 1945. These tables are presented in the following form. The primary division is by major geographical region (North America, Latin America, Europe, Middle East, South Asia, Far East, Oceania, Sub-Saharan Africa, and Other Areas). Each of these regions is further subdivided into nations (e.g., Burma, Cambodia, China, Indonesia, Japan, Korea, Laos, Malaysia, Mongolia, Philippines, Taiwan, Tibet, and Vietnam). Within each nation, the following information is given on each conflict listed: inclusive dates (e.g., 1952-1963), a short description (e.g., "Independence from UK"), and the estimated number of civilian and military deaths (together with their total). Deaths due to massacres, political violence, and famines associated with the conflicts are included. Essentially no situational descriptors are included, apart from the general location, time span, and such extremely succinct descriptions of the conflict as "Independence from France," "Bosnia vs. Turkey," and the like.

The data in the 1988 edition are based on a compilation made by Dr. William Eckhardt. His notes observe that:

- (1) The bigger the war, and the more recent, the less reliable are the estimates.
- (2) Most of the war deaths of three centuries have occurred in the 20th, which has so far accounted for over 90 percent of the deaths in wars since 1700 AD.
- (3) Starters of wars are infrequently the winners. In the 20th Century, starters on average have won 39 percent of the wars.

(4) The geography of warfare has changed radically. Europe was the principal site of wars and war deaths over the three-century span, but since World War II, only one conflict (Hungary) has taken place in Europe. All other wars have been fought in the Third World. They have not, however, been without involvement of the major powers, and recently this involvement, often indirect and covert, appears to have increased.

(5) As for "causes" of wars, the most frequent objectives were for territory or independence. However, civil wars, representing power conflicts within nations, have increased sharply in the 20th Century, and are now by far the major form of warfare.

We elected to omit these data from our catalog of data bases because the material appears to fall under protection of the copyright laws, and because its validity and relevance to attrition rates in battle seems somewhat questionable. In particular, it is difficult to discern exactly what sources of information were used in compiling these data.

1. The Human Relations Area Files are maintained by the Human Relations Area Files, Incorporated (HRAF), a nonprofit research and educational organization centered at Yale University in New Haven, Connecticut. They can be contacted at P.O. Box 2054 Yale Station, New Haven, CT 06520-2054, phone: 203-777-2334. Since 1949 HRAF has devoted its resources to developing programs and services to encourage and facilitate the worldwide comparative study of culture, society, and human behavior. Today, the HRAF Archive, with some 3,500,000 pages of information organized into files on more than 330 different cultural groups and available on over 21,500 microfiche, considers itself to be the world's most complete source of information on the ways of life of people around the world. Today, over 80 percent of the major research universities in the United States are members of HRAF and maintain the complete set of the HRAF Archive. The coverage is primarily of preindustrial, or even primitive, cultures. It consists of a collection of mostly primary descriptive materials (books, articles, manuscripts, etc.) classified both by the culture or society to which they pertain and the topics discussed therein. Thus, the Archive combines the basic elements of a test archive with those of an information retrieval system for easy user access and rapid information retrieval. Every year for the past 38 years the Archive has been updated and expanded to broaden coverage of the cultures of the world. The collection is organized, first, into culture files, with each culture file containing information on a particular cultural group such as the Zuni Indians, the Rural Irish, the Yoruba of Nigeria, or the Central Thai. The full collection contains files on more than 330 different cultural groups. The collection is organized, second, by subject, according to the more than 700 subject categories in the *Outline of Cultural Materials (OCM)*. Thus, all information on a given subject is brought together in one place in each cultural file, making for rapid and comprehensive information retrieval. All materials are provided on silver halide microfiche. The 1991 Annual Report noted that a program to computerize major portions of the Archive had been initiated.

Dr. Melvin Ember, President of Human Relations Area Files, Inc., delivered a paper on "Fear of Disasters as an Engine of History: Resource Crises, Warfare, and Interpersonal Aggression" to the multidisciplinary conference on "What is the Engine of History?" at Texas A&M University, 27-29 October 1988. This paper, based on HRAF data, makes the following points.

(1) Contrary to what many have wanted to believe, most societies known to anthropology have had warfare, and the warfare probably occurred a lot more often than even we are used to in the modern world. In the societies we have looked at that were described before pacification, nearly 75 percent had warfare at least once every 2 years.

(2) In the few cases where we have detailed information on the number of people killed over time, it seems that "primitive" warfare might have been even more lethal *proportionately* than modern warfare. For example, the Mae Enga of the New Guinea Highlands lost about 25 percent of their males because of intervillage warfare, and other societies in New Guinea and Amazonia lost between 19.5 and 30 percent.

(3) Our results strongly suggest that people in preindustrial societies mostly go to war to moderate or cushion the impact of expected but unpredictable disasters that destroy food resources—disasters such as droughts, floods, storms, killing frosts, and locust infestations.

(4) The notion that greater social complexity increases the likelihood of war is only very weakly supported by the data.

(5) Chronic scarcity does not seem to increase the likelihood of warfare.

(6) Our results do not support the view that punitive or nonrewarding socialization has a major effect on the likelihood of war. Indeed, our only near-significant result suggests that indulgence in infancy may increase rather than decrease the likelihood of war.

(7) The notion that sexual frustration increases the likelihood of war is not supported by the data.

(8) Societies that encourage boys to be aggressive have more warfare, but this is interpreted as a consequence of frequent wars rather than a cause of them.

(9) The notion that a higher level of interpersonal aggression increases the frequency of warfare is only weakly supported by the data. As of now, we suspect that higher aggression frequencies are mostly indirect consequences of a high frequency of warfare. High frequency of warfare (induced by disasters) induces the culture to produce fierce warriors, and this greater aggressiveness spills over into higher rates of homicide and other forms of interpersonal violence within the culture. Also, a high frequency of warfare may "legitimize" violence (another study has found that homicide rates increase after episodes of war). Moreover, the threat of disasters may have a direct effect on aggressive behavior

by creating a great deal of free-floating anxiety, insecurity, and a feeling of lack of control (another study has found that capricious violence in folktales is associated with a high threat of natural disasters).

The authors themselves caution that the research reported here is based on preindustrial cultures and may not apply to modern industrialized societies with their complex international dependencies. This data base was omitted from the catalog because it is proprietary, not reduced to a form usable with personal computers, and because it is of doubtful relevance to personnel attrition rates in land combat operations.

m. Historical Evaluation and Research Organization (HERO) Report "Modern Artillery Experience in Combat: Evaluation of Historical Artillery Combat Activities," in two volumes, prepared for the US Army Field Artillery School under Contract No. DABT58-82-C0069, 30 September 1983, UNCLASSIFIED. This study contains numerous tabulations on various aspects of artillery operations in World War II, mainly from North Africa and Italy.

n. Federal Republic of Germany Armed Forces Research and Training Office Report "Information on Consumption of Ammunition by Land Forces Since 1939," Bergisch Gladbach 1, June 1986, FOR RELEASE ONLY WITH AUTHORIZATION OF FED. MIN. OF DEF. Contact the controlling office cited above for further information about this report.

o. "Study of Land/Air Trade-Offs (SLAT), Volume VII, Historical Data Base," Study 64, Center for Naval Analyses, Naval Warfare Analysis Group, L. Feldman, *et al.*, undated (but circa 1970), 100 pp, UNCLASSIFIED, available from DTIC (AD-869479). This contains the historical data on United Nations operations in Korea that were simulated during the SLAT studies. However, much of the tabulated data is illegible in the copies made available by DTIC. In addition, the subject document states (p A-4) that "It must be strongly emphasized that the quality of the data is rather poor. Enemy casualties are estimates and probably inflated. Enemy strengths are essentially unknown, unless we make some rather broad assumptions. There are internal inconsistencies in the data, such as on May 7 and 8 when the 2nd Infantry Division reported no sorties by air but 95 and 85 enemy casualties due to air, respectively. Some of the information is presented corps-wide, and we have split this up somewhat arbitrarily. The positions on the maps are of unknown accuracy." The questionable validity of the data, together with the impracticability of deciphering illegible tabulations, led us to decide not to include these data.

p. Fox, William F., *Regimental Losses in the American Civil War, 1861-1865*, Albany Publishing Company, Albany, NY, 1889, 582 pp, UNCLASSIFIED, available from libraries. Although not included in the catalog of data bases because its material is too voluminous to digitize within the scope of this effort, this source records various statistics about losses to regimental-sized units during the Civil War, mostly on the Union side.

q. Love, MAJ Albert G., *Medical and Casualty Statistics*, US Army Surgeon General's Office, Volume XV (Statistics), part 2 (Medical and Casualty Statistics), of the series *The Medical Department of the United States Army in the World War* [i.e., World War I], published by the US Government Printing Office, Washington, DC, 1925, 1,368 pp, UNCLASSIFIED, available from US Army Military History Institute (UM24 1917-18 A45 1921 V.15 pt. 2). (Part 1 of Volume XV, entitled "Army Anthropology. Based on Observations Made on Draft Recruits, 1917-1918, and on Veterans at Demobilization, 1919," is not relevant to attrition rate analysis.) Although not included in the catalog of data bases because its material is too voluminous to reduce to digital form within the scope of this effort, this document contains a comprehensive account of US Army casualties in World War I.

r. Mitchell, T. J., and Smith, G. M., *Casualties and Medical Statistics of the Great War*, His Majesty's Stationery Office, number 57-484, 1931, 382 pp, UNCLASSIFIED, available from libraries (copy held by US Army Command and General Staff College, Leavenworth, KS, under call number 940.4754 1 M 682m). This is the final volume of the *Official Medical History of the War*, which deals with the statistical aspect of casualties. Although not included in the catalog of data bases because its material is too voluminous to reduce to digital form within the scope of this effort, this volume summarizes information from the records on over 11 million casualties sustained by the British Expeditionary Forces at home and in the various campaigns overseas during the Great War. In that total are included both the casualties on the battle field and those occasioned by disease and injury.

s. Otis, George A.; and Huntington, D. L., *Surgical History*, (Vol II, Pt III, Medical and Surgical History of the War of the Rebellion), Office of The Surgeon General, US Army, 1883, 997 pp, UNCLASSIFIED, available from libraries (copy held by University Library of North Dakota as 614.0973 U58m, v2, pt 3). This is the concluding volume in the series "The Medical and Surgical History of the War of the Rebellion," and its Chapter XII is "devoted to a general consideration of wounds and complications, in the course of which the endeavor will be made to use, to the best advantage, the large amount of material accumulated during the late war ...". Although not included in the catalog of data bases because its material is too voluminous to reduce to digital form within the scope of this effort, this is a most careful and authoritative analysis of wounded in the American Civil War.

t. Reister, Frank A., *Medical Statistics in World War II*, Office of The Surgeon General, US Army, Library of Congress Catalog Card Number: 75-600004, 1975, 1,215 pp, UNCLASSIFIED, available from libraries. Although not included in the catalog of data bases because its material is too voluminous to reduce to digital form within the scope of this effort, this is an official US Army study of medical statistics from World War II.

u. Reister, Frank A., *Battle Casualties and Medical Statistics, US Army Experience in the Korean War*, The Surgeon General, Department of the Army, 1986, 172 pp, UNCLASSIFIED, available from libraries. First printed in 1973 as Center of Military History Publication CMH Pub 82-5. These data have been assembled with the planner in mind, based on a knowledge of material considered essential by systems analysts engaged in combat development, wargaming, weapons evaluation, medicological problems, and in medical research concerning battle casualties. Although not included in the catalog of data bases because its material is too voluminous to reduce to digital form within the scope of this effort, this volume is an official source of tabulated casualty data for US forces in the Korean War and is one of the major classics in this field.

D-3. OTHER SOURCES OF INFORMATION. The reader should be aware of the following additional sources of information on losses.

a. **DTIC.** Many works are available from the Defense Technical Information Center (DTIC). Your librarian should know how to use the DTIC document accession number (the so-called AD-number) to obtain documents from DTIC. If needed, further information on DTIC may be obtained from:

Defense Logistics Agency
Defense Technical Information Center (DTIC)
Bldg 5, Cameron Station
5010 Duke Street, Alexandria, VA 22314-6145
Document Information, 703-274-7633.
Answering Service for User Requests, 703-274-6811.

b. **NTIS.** Those who have no access to DTIC can usually obtain unclassified AD documents from the National Technical Information Service (NTIS). For information on that procedure, contact:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151
Document Sales Desk, 703-487-4650.
General Information, 703-487-4600.
Computer Products, 703-487-4763.
Customer Services, 703-487-4660.

c. **PASBA.** Various organizations are tasked with the responsibility for collecting, processing, analyzing, and disseminating selected types of casualty data. One of these is PASBA:

US Army
Patient Administration Systems and Biostatistics Activity (PASBA)
Fort Sam Houston, Texas 78234-6070.

Its mission is to "implement policy and provide guidance on patient administration and operate a worldwide medical data collection system in support of United States Army Health Services

Command." As of late 1992, it consisted of approximately 92 military and civilian personnel with the following organization and functions.

(1) **Office of the Commander.** Provides consultative service to the Office of the Army Surgeon General; HQ 7th Medical Command, Europe; HQ 18th Medical Command, Korea; and Department of Army pertaining to patient administration.

(2) **Chief Statistician.** Develops organizational statistical policy and provides guidance on statistical methodology. Monitors biostatistical projects for the activity. Reviews all studies which have statistical analyses prior to release. Performs liaison duties with Department of Army, federal, state, and private agencies for development of programs and exchange of data. Serves as staff advisor to commander on matters pertaining to statistical methods and policy.

(3) **Medical Records Consultant.** Implements policy and provides guidance on medical records administration. Provides consultation services to US Army Health Services Command (HSC) staff elements and its hospitals on medical records administration to include technical assistance and staff visitation regarding functional organization, chart analysis, diagnostic coding, standard nomenclature, preparation of coding transcripts, file management, medical records research, development of medical records forms, standardization of equipment for medical records forms, standardization of equipment for medical records activities, requirements of various national health and accrediting agencies, and Joint Commission for Accreditation of Healthcare Organizations (JCAHO). Develops policies and procedures to improve the efficiency of inpatient, outpatient, and health records administration and related systems, and monitors their implementation within HSC.

(4) **Administration Division.** Provides administration, logistical, and information management support to the activity's technical divisions.

(5) **Biostatistics Division.** Provides statistical analysis in support of the Army Medical Department (AMEDD). Receives, interprets, and provides medical statistical data for varied users. Data is used by the AMEDD in making personnel and financial resourcing decisions. The information is also utilized for clinical research and statistical studies. The division maintains a computerized data base for patients treated in Army hospitals. And, as such, plays a key role in the development of statistical models and automation systems.

(6) **Patient Administration Systems Division.** Operates the Individual Patient Data System (IPDS) and the central registry for the Army Family Advocacy Program System (AFAPS). The IPDS is an automated worldwide medical data collection system. Data received is screened for completeness and validity before being made available for the AMEDD and Department of Army users. The division also maintains the automated worldwide AFAPS central registry containing data on child and spouse abuse.

(7) **Patient Administration Operations Division.** Implements policies and provides guidance to Army medical treatment facilities (MTF) on patient administration. The division provides technical guidance, training, and assistance on matters pertaining to patient administration management in such areas as medical eligibility, entitlements, business office operations, casualty reporting, decedent affairs, medical disability, procurement of civilian medical services, and sharing of facilities with other federal medical facilities.

(8) Commercial phone numbers are of the form 512-212-####; DSN numbers of the form 471-####. Points of contact with the phone digits ####, as of late 1992, were as follows:

Office of the Commander (HSHI-QZ)	1102
Administration Division (HSHI-QA)	0780
Biostatistics Division (HSHI-QB)	0688
Patient Administration Systems Division (HSHI-QP)	0797
Patient Administration Operations Division (HSHI-QR)	2978
Chief Statistician (HSHI-QZ)	0471
Medical Records Consultant (HSHI-QZ)	0471

d. **SURVIAC.** Another organization tasked with the responsibility for collecting, processing, analyzing, and disseminating selected types of loss data is SURVIAC:

US Department of Defense
Survivability/Vulnerability Information Analysis Center (SURVIAC)
WL/FIVS/SURVIAC
Wright-Patterson AFB, Ohio 45433-6553

Its mission is to provide an information center on nonnuclear survivability and lethality issues, focused chiefly on aircraft and other vehicle losses. It maintains computerized data bases and computer-indexed referenced libraries. In addition to its own holdings, it can access other existing data bases such as DTIC, the NASA Scientific and Technical Information Library, Lessons Learned Program, and the Chemical Defense Data Base. As a full-service information analysis center, SURVIAC identifies, reviews, evaluates, and stores relevant scientific and technical data on all aspects of nonnuclear survivability and lethality. SURVIAC determines and responds to user community needs and provides technical advice and support in such areas as survivability design, key technologies, survivability assessments, applied methodologies, and munitions effectiveness analysis. Technical areas of interest extend to:

- Weapon system, subsystem, and component physical and functional descriptions and characteristics.
- Vulnerability, vulnerability reduction, susceptibility, and susceptibility reduction assessments.
- Live fire test data and analyses.
- Survivability, lethality, and munitions effectiveness assessments.
- Combat systems operations, damage, and repair.
- System signatures, target detection, acquisition, and tracking.

- Threat launch, flyout, and fuzing characteristics.
- Countermeasures, counter-countermeasures, and threat effects.

SURVIAC's major reference libraries include:

- Survivability/Vulnerability Reference Library (reports, studies, analyses, and raw data).
- Laser Reference Library (test, research and development, and directed energy weapons).
- Aircraft Battle Damage Repair Library (reports and data on aircraft battle damage repair).
- Vehicle Signatures Library (reports and data on vehicle signatures).
- Combat Evaluation Library (reports, studies, and analyses from US Army experience in Southeast Asia).

SURVIAC's data base services include:

- ACFTDAB (Southeast Asia fixed wing aircraft).
- HELODAB (Southeast Asia rotary wing aircraft).
- GNVEHSEADB (Southeast Asia ground vehicle).
- LASERDAB (laser shots against a variety of materials).
- RAM TEAM (rapid area maintenance team data base).
- Yom Kippur (Arab-Israeli tank and personnel carrier; fixed wing aircraft).
- JUST CAUSE (battle damage repair from 1989 Panama incident).
- Pk/h Matrix (critical component and supporting test data requirements matrix).

SURVIAC's central office can be reached on DSN 785-7840 or commercial 513-255-4840. Its Washington, DC area satellite office is located at Booz-Allen & Hamilton, Inc., 4001 N. Fairfax Drive, Suite 650, Arlington, VA 22203 and can be reached on 703-528-8080.

(THIS PAGE INTENTIONALLY LEFT BLANK)

APPENDIX E DISTRIBUTION

Addressee	No. of copies
Deputy Under Secretary of the Army (Operations Research), Room 2E660 The Pentagon Washington, DC 20310	1
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZXA Washington, DC 20310	1
Deputy Chief of Staff for Logistics Headquarters, Department of the Army ATTN: DALO-ZXA-A Room 3D572, The Pentagon Washington, DC 20310-0580	1
Commander Combined Arms Support Command ATTN: ATCL-CFC Fort Lee, VA 23801-6000	1
Office of the Secretary of the Army Correspondence & Records Center Management Systems & Support ATTN: JDMSS-CRC ROOM 3D679, The Pentagon Washington, DC 20310-0105	1
Office of The Surgeon General ATTN: DASG-HCD 5109 Leesburg Pike Falls Church, VA 22041-3258	1
Director US Army TRADOC Analysis Command-WSMR ATTN: ATRC-WSL White Sands Missile Range, NM 88002-5502	1
Commander, TRAC ATTN: ATRC-TD Fort Leavenworth, KS 66027-5200	1
HQ TRADOC ATTN: ATAN-S Fort Monroe, VA 23651-5143	1

Addressee	No. of copies
Director US Army Materiel Systems Analysis Activity ATTN: AMXSJ-LM Aberdeen Proving Ground, MD 21005-5071	1
Director US Army Research Laboratories ATTN: AMSRL-SL-B (Klopcc) Aberdeen Proving Ground, MD 21005-5068	1
Commander Army Research Institute ATTN: PERI-SM 5001 Eisenhower Avenue Alexandria, VA 22333-5600	1
Commander US Total Army Personnel Agency ATTN: DAPC-ZA 200 Stovall Street Alexandria, VA 22332	1
Defense Technical Information Center ATTN: DTIC-FDAC Cameron Station Alexandria, VA 22314-6145	2
USASCAF The Pentagon Library ATTN: JDHQ-LR (Army Studies) Room 1A518, The Pentagon Washington, DC 20310-6000	1
Commandant US Army War College Operations Group ATTN: AWCM-A Carlisle Barracks, PA 17013-5050	1
Air War College ATTN: AU/CADRE/WGOI Maxwell Air Force Base, AL 36112-5532	1
President US Navy War College ATTN: Code E-111/Library Newport, RI 02841-5010	1

Addressee	No. of copies
President National Defense University ATTN: NDU-LD-CDC Washington, DC 20319-6000	1
Commandant Armed Forces Staff College ATTN: Library Room B-201 7800 Hampton Blvd Norfolk, VA 23511-6097	1
Commandant US Army Command and General Staff College ATTN: ATZL-SWS-L (Mail) Fort Leavenworth, KS 66027-6900	1
Superintendent Naval Postgraduate School ATTN: Security Manager Monterey, CA 93940	1
Commander US Army Health Services Command ATTN: HSOP-FSI Fort Sam Houston, TX 78234-6000	1
Commander US Army Medical Research and Development Command ATTN: SGRD-OP (Mr. Adams) Fort Detrick, MD 21701	1
Director Strategic Studies Institute ATTN: AWC Carlisle Barracks, PA 17013-5050	1
Commander/Director US Army Engineer Studies Center Casey Building, No. 2594 ATTN: ESC-AO (Security Officer) Fort Belvoir, VA 22060-5583	1
Commander US Army Training and Doctrine Command ATTN: ATIM-OPM Fort Monroe, VA 23651-5000	1

Addressee	No. of copies
Headquarters Tactical Air Command ATTN: DAAS Langley Air Force Base, VA 23665-5001	1
Commandant Air Force Institute of Technology ATTN: AFIT-EN Wright-Patterson AFB, OH 45433	1
Chief of Naval Operations ATTN: OP-09B34F1 Room 4C479, The Pentagon Washington, DC 20350	1
United States General Accounting Office (GAO) National Security and International Affairs Division Washington, DC 20548	1
Director Program Analysis and Evaluation (PA&E) Office of the Secretary of Defense Room 2E330 The Pentagon, Washington, DC 20310	1
Casualty Care Research Center Dept. of Military Medicine Uniformed Service University of Health Services 4301 Jones Bridge Rd. Bethesda, MD 20814	1
The Borden Institute Textbook of Military Medicine Walter Reed Army Medical Center Building 1, Room D-100 Washington DC 20307-5001	1
US Army Intelligence and Threat Analysis Center ATTN: IAIT-RAS (Cocca) Building 213 Washington Naval Yard Washington, DC 20374-5085	1
Logistics Management Institute (LMI) ATTN: Mr. George Kuhn 6400 Goldsboro Road Bethesda, MD 20817-5886	1

Addressee	No. of copies
-----------	------------------

The RAND Corporation
ATTN: Eric Larson
1700 Main Street
PO Box 2138
Santa Monica, CA 90406-2138

1

Internal Distribution:

Reference copy:
Unclassified Library

2

Record copy:
Originating office (CSCA-MV)

1

Other copies:
Dr. Robert L. Helmbold

20

Addressee	No. of copies
-----------	------------------

(THIS PAGE INTENTIONALLY LEFT BLANK)

GLOSSARY

GLOSSARY-1. Some of the abbreviations and special terms used in this document are listed below. If the definition given is an official one, the organizations that have adopted it are given in parentheses; otherwise no indication of its adoption is given. Note that the definitions used by other countries or by the US in earlier times may differ more or less from those given below, and may be interpreted in various ways even within the US Department of Defense.

GLOSSARY-2. Definitions of terms and abbreviations:

Battle casualty: (DOD) Any casualty incurred in action. "In action" characterizes the casualty status as having been the direct result of hostile action, sustained in combat or relating thereto, or sustained going to or returning from a combat mission provided that the occurrence was directly related to hostile action. Included are persons killed or wounded mistakenly or accidentally by friendly fire directed at a hostile force or what is thought to be a hostile force. However, not to be considered as sustained in action and thereby not to be interpreted as battle casualties are injuries due to the elements, self-inflicted wounds, and, except in unusual cases, wounds or death inflicted by friendly forces while the individual is in absent without leave or dropped from rolls status or is voluntarily absent from a place of duty. See also died of wounds received in action; nonbattle casualty; wounded.

Bloody losses: The sum of the KIA and WIA.

Casualty: (DOD, IADB) Any person who is lost to the organization by reason of having been declared dead, wounded, injured, diseased, interned, captured, retained, missing, missing in action, beleaguered, besieged or detained; see also battle casualty; nonbattle casualty; wounded.

CMIA: Captured or missing in action. See POW and MIA.

CRO: Carded for record only. (Adapted from Beebe, Gilbert W.; and De Bakey, Michael E., "Battle Casualties: Incidence, Mortality, and Logistic Considerations," Charles C. Thomas (publisher), 1952). Basically, admissions to a medical treatment facility include all cases admitted for medical care and not returned to duty on the same calendar day as that on which first seen, but they also include certain other cases treated on an outpatient (duty) status, designated as carded for record only.

DNBI: Disease and nonbattle injury. Personnel treated for diseases and for injuries not received in action. See Nonbattle casualty.

DOW: Died of wounds received in action (DOD, NATO). A battle casualty who dies of wound or other injuries received in action, after having reached a medical treatment facility. See also killed in action.

DTIC: Defense Technical Information Center.

KIA: Killed in action (DOD, NATO, IADB). A battle casualty who is killed outright or who dies as a result of wounds or other injuries before reaching a medical treatment facility. See also died of wounds received in action.

Losses: (Adapted from FM 101-10-1/2, Staff Officers' Field Manual, *Organizational, Technical, and Logistical Data Planning Factors*, October 1987). A personnel loss is any reduction in the assigned strength of a unit. Personnel losses are recorded in three general categories: battle, nonbattle, and administrative.

- **Battle losses** are those incurred in action. They include wounded or injured in action (including those who died of wounds and died of injuries received in action), killed in action, and missing in action or captured by the enemy.

- **Nonbattle losses** are those not directly attributable to action regardless of when sustained. They include nonbattle dead, nonbattle accident/injury, nonbattle missing, and illness/disease.

- **Administrative losses** are those resulting from transfer from the unit, absence without leave, desertion, personnel rotation, and discharges.

LWIA: Lightly wounded in action (*cf.* Slightly wounded).

MIA: (adapted from FM 101-10-1/2, Staff Officers' Field Manual, *Organizational, Technical, and Logistical Data Planning Factors*, October 1987). Missing in action describes battle casualties whose whereabouts or fate cannot be determined and who are not known to be in an unauthorized absence status (desertion or absence without leave). Missing in action casualties are not usually included in medical statistical records or reports received by The Surgeon General, but are reportable to The Adjutant General.

Nonbattle casualty: (DOD, NATO, IADB) A person who is not a battle casualty, but who is lost to his organization by reason of disease or injury, including persons dying from disease or injury, or by reason of being missing where the absence does not appear to be voluntary or due to enemy action. See also battle casualty; wounded.

Nonbloody loss: Battle casualties other than KIA and WIA; includes (for example) MIA, POW, absent without leave, stragglers, and deserters.

NP: Neuropsychiatric.

POW: Prisoner of war. Detainee (DOD). A term used to refer to any person captured or otherwise detained by an armed force. (According to FM 101-10-1/2, Staff Officers' Field Manual, *Organizational, Technical, and Logistical Data Planning Factors*, October 1987, captured describes all battle casualties known to have been taken into custody by a hostile force as a result of and for reasons arising out of any armed conflict in which US armed forces are engaged. Captured casualties are not usually included in medical statistical records or reports received by The Surgeon General but are reported to The Adjutant General.)

Seriously wounded: (DOD, IADB) A stretcher case. See also WIA.

Slightly wounded: (DOD, IADB) A casualty that is a sitting or walking case. See also WIA.

SWIA: Seriously wounded in action (*cf.* Seriously wounded).

TBC: Total battle casualties (the sum of the KIA, WIA, MIA, and CMIA).

WIA: Wounded in action (DOD, NATO, IADB). A battle casualty other than "killed in action" who has incurred an injury due to an external agent or cause. The term encompasses all kinds of wounds and other injuries incurred in action, whether there is a piercing of the body, as in a penetrating or perforated wound, or none, as in the contused wound; all fractures, burns, blast concussions, all effects of biological and chemical warfare agents, the effects of exposure to ionizing radiation, or any other destructive weapon or agent.